(20)

SPERRY UNIVAC
COMPUTER SYSTEMS

ADA036187

PROSODIC AIDS TO SPEECH RECOGNITION:

IX. ACOUSTIC-PROSODIC PATTERNS IN SELECTED ENGLISH PHRASE STRUCTURES

by

WAYNE A. LEA

DEFENSE SYSTEMS DIVISION
ST. PAUL, MINNESOTA
(612) 456-2434

Final Technical Report Submitted To:

ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22209

Attention: DIRECTOR, IPTO

31 December 1976

Report No. PX11963

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC 15-73-C-0310, ARPA Order No. 2010. The views and conclusions contained in this document are those of the author, and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

SPERRY UNIVAC
COMPUTER SYSTEMS

PROSODIC AIDS TO SPEECH RECOGNITION:

IX. ACOUSTIC-PROSODIC PATTERNS IN SELECTED ENGLISH PHRASE STRUCTURES

See 1473

by

WAYNE A. LEA

DEFENSE SYSTEMS DIVISION
ST. PAUL, MINNESOTA
(612) 456-2434

Final Technical Report Submitted To:

ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22209

Attention: DIRECTOR, IPTO

31 December 1976

Report No. PX 11963

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC 15-73-C-0310, ARPA Order No. 2010. The views and conclusions contained in this document are those of the author, and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

PREFACE

This is the ninth, and last, in a series of reports on Prosodic Aids to Speech Recognition. The previous reports appeared as follows:

I.	Basic Algorithms and Stress Studies	1 October, 1972	PX 7940
II.	Syntactic Segmentation and Stressed Syllable Location	15 April, 1973	PX 10232
III.	Relationships Between Stress and Phonemic Recognition Results	21 September, 1973	PX 10430
IV.	A General Strategy for Prosodically-Guided Speech Understanding	29 March, 1974	PX 10791
V.	A Summary of Results to Date	31 October, 1974	PX 11087
VI.	Timing Cues to Linguistic Structure and Improved Computer Programs	31 March, 1975	PX 11239
VII.	Experiments on Detecting and Locating Phrase Boundaries	14 November, 1975	PX 11534
VIII.	Listeners' Perceptions of Selected English Stress Patterns	21 June 1976	PX 11711

This research was supported by the Advanced Research Projects Agency of the Department of Defense, under Contract No. DAHC15-73-C-0310, ARPA Order No. 2010.

I would like to acknowledge the help during various phases of our project of the following other Sperry Univac researchers:

Dr. Mark Medress
Dr. Timothy Diller,
Dean Kloker, and
Toby Skinner.

SUMMARY

For four and one half years, Sperry Univac has been engaged in research on prosodic structures and their use in speech understanding systems, as a part of the large ARPA Speech Understanding Research (SUR) program. This, the final report on the effort, describes our two most recent studies and summarizes the results of our entire work for ARPA.

One of our two major accomplishments during this final reporting period was to complete our aspects of a cooperative study with Bolt Beranek and Newman (BBN), to develop prosodic aids for the parser in the BBN "Hear What I Mean" (HWIM) speech understanding system. We marked every transition arc in the SMALLGRAM grammar used in the HWIM system, indicating whether the word associated with that point in generated syntactic structures would or would not be immediately preceded by a phrase boundary detected from the fundamental frequency (Fo) contours associated with the speaking of that structure. Previous research had shown that the first stressed syllable of major syntactic phrases in spoken sentences would be immediately preceded by a fall-rise "valley" in Fo, so that the first stressed words of a phrase should be marked in the grammar as expected to be preceded by a boundary.

We proposed to BBN that procedures be implemented that would increase the score on words hypothesized by the HWIM system if those words were expected to be preceded by a boundary and application of the Fo-based phrase boundary detector showed a boundary occurring just before the hypothesized position of that word in the utterance. Thus if a boundary was detected, those words were rewarded (increased in score) that predicted a boundary at that point. If a boundary was expected before a hypothesized word (that is, the word was marked in the grammar as one expected to be immediately preceded by a boundary), but no boundary was detected in the Fo data, the score on that hypothesized word would be reduced substantially. In essence, if you expect a boundary, you better get it, and if you do, you should try that word or phrase hypothesis earlier than ones whose expected prosodies disagree with the detected prosody.

To test this specific concept of prosodic aids to parsing, we processed a total of sixteen sentences through our prosodic analysis tools, and compared the detected phrase boundaries with BBN's control and parsing traces of HWIM's attempts to understand those sentences. BBN's latest method of hypothesis scoring, based on a shortfall density function, had to be accomodated. Also, we needed to help BBN define exactly when a boundary may be said to "immediately precede" a hypothesized word. We found that, for all the sentences, the prosodic adjustments in scores on one-word and multiple-word theories would help correct word sequences to be tried before erroneous ones. Thus, prosodics clearly seemed to offer frequent chances of helping the parser.

BBN consequently implemented procedures for using the marked arcs of the grammar, the Fo-detected phrase boundaries, and the adjustment of scores on hypothesized word sequences or "theories". Unfortunately their ARPA project ended before these procedures could be tested to confirm the value of the prosodic aids to parsing. Still, we believe that our hand analyses showed real promise of intonational phrase boundaries guiding the parser to correct parses.

One other major accomplishment during these last few months was to analyze the acoustic prosodic patterns in 255 of our carefully-designed data base sentences. Perceived stress patterns for those sentences had already been obtained in our previous studies. We found that over 91% of the syllables were correctly detected, despite the difficult all-sonorant sequences in many of these sentences. Also, 76% of the expected phrase boundaries were correctly detected, and 92% of the perceived stresses were correctly located. These were very encouraging results. However, also of interest were the regularities in boundary placement, stress assignment, and intonation found from these studies of sentences with minimally-contrastive structures.

We found that, following a phrase with a stress, boundaries occurred before the first stresses of: noun phrases; sentence adverbs; conjuncts, relative clauses; parantheticals; main verbs; and auxiliaries with stresses (e.g., containing a negative). Boundaries also occur before stressed prepositions and between the parts of a compound noun. Boundaries were

more reliably detected if unstressed syllables intervened between the last stress of the previous phrase and the first stress of the following phrase. Unexpected boundaries regularly occurred before (unstressed) relative pronouns, and after the inverted auxiliary in yes/no questions (before NP subjects or, if the subject was a pronoun, before the main verb). About 8% of all detected boundaries were false, caused by Fo variations near obstruents and utterance -initial and final variations in Fo that can be ruled out as wrong places for boundaries.

We learned from those controlled studies that 99% of all first stresses in sentences occur just before the Fo peak of the sentence. The stress location algorithm can be significantly improved by incorporating this fact. Many errors (missed stresses and unstresses falsely called stresses) resulted from failures to separate two short syllables, so they look like one long ("stressed") syllable. The test for prepausal stresses also needs some refinement. While there is room for improvement in the stress location program, it in general is giving good results.

An overview of all our work for ARPA over the last four and one half years is given in section 4 of this report. In brief, we have provided extensive evidence for the value of prosodics in speech understanding systems, have cooperated in many common tasks of the total SUR program, have provided programs and ideas for the use of prosodics in speech understanding systems, and have conducted a series of experiments to learn regularities about the whole gamut of prosodic structures. Rather than reiterate that extensive list of contributions here, I recommend the reader read Table VII (page 41 to 43) at this time. It presents most of our major accomplishments in an organized manner.

While the experimental results from our studies are many, the major conclusions are simply stated. We have shown that prosodic features can be effectively used to locate regions of phonetic reliability, guide phonological analyses, aid word matching, and provide parsing and higher level linguistic analyzers with independent information about the structures of spoken sentences. Prosodics show promise of providing major improvements in speech understanding, but no complete test of their effectiveness has been accomplished, and much is yet to be learned about the prosodic regularities of English that

may aid speech understanding. Further work is needed, and we have provided the analysis tools, experimental designs, carefully designed speech databases, hypotheses, and program plans that could make such work possible. There is no question but that prosodic analysis should play an important role in future speech understanding systems.

TABLE OF CONTENTS

1. INTRODUCTION

This is the nineth, and last, in a series of reports on Sperry Univac's contracts under the Speech Understanding Research program sponsored by the Advanced Research Project Agency (ARPA). Sperry Univac's research is concerned with extracting prosodic information from the acoustic waveform of connected speech (sentences and discourses), and using that prosodic information to detect phrase boundaries, locate stressed syllables, determine rhythm and rate of speech, and apply such prosodic features to guiding word matching, syntactic parsing, and semantic analyses.

Sperry Univac's research has been basically two-pronged: (1) developing prosodic analysis tools and providing other services to speech understanding systems builders; and (2) conducting experiments suitable for determining exactly how prosodic patterns relate to sentence structures. The experiments have a definite practical goal in mind, however. They are intended to provide adequate understanding so that prosodic tools can be implemented and improved in such a way as to provide substantial aids to other aspects of the speech understanding process.

In Section 4 of this report, the reader will find a review of all the research done by Sperry Univac to provide prosodic aids to speech understanding and conduct research on prosodic structures. Before we get to that review, however, we must first consider the final two experimental studies just completed in this program. In this last half year, we have completed our study of how intonationally-detected phrase boundaries might be used in the Bolt Beranek and Newman HWIM ("Hear What I Mean") system. As described in Section 2, the most recent procedures for scoring alternative syntactic theories in the HWIM system were considered, the latest form of the grammar was marked for prosodic expectations, and a set of ten additional sentences were processed to determine whether detected phrase boundaries could help the HWIM parser. In the last days of their SUR project, BBN was not able to test our ideas as we had hoped, but our analyses do suggest good potential for prosodic information to aid in obtaining correct parses of sentences.

In Section 3, we report on the acoustic-prosodic patterns found in 255 representative sentences taken from our 3300-sentence speech data base. With these carefully controlled sentence structures, we were able to determine various

prosodic regularities that accompany various English phrase structures. Several questions still remain, however, and further tests are needed.

Following the review in section 4, section 5 covers our general conclusions and recommendations for further work. Our recent studies of prosodic patterns in controlled English texts have shown the great value of the speech data base and the need to test a variety of hypotheses about prosodic patterns. Our applications of prosodic analysis tools to guiding speech understanding systems have shown the need for further research on the direct use of prosodic information in speech understanding systems. Ideas for furthering such work are suggested.

References in section 6 are followed by a listing in Appendix A of the sentences and their prosodic structures, and, in Appendix B, a listing of all the presentations and publications completed at Sperry Univac within the ARPA SUR program.

2. PROSODIC AIDS FOR THE BBN PARSER

2.1 Previous Results

In our previous semiannual report (Lea, 1976c), we reported on an initial experiment to develop means of using prosodic information to help guide the parsing and control components of the BBN HWIM System. This study was an attempt to show on paper that prosodics can be of some use, even before a specific procedure is coded and tested within the HWIM system. It appeared that some aid can be provided, but the ultimate success, in obtaining more correct parses or more rapid and efficient parses, awaited further testing and implementation, which we hoped to accomplish in this last six months, with BBN's cooperation. Let me first summarize where we were, and then outline our further work. I will follow the form of presentation used in describing this work at the 92nd Meeting of the Acoustical Society of America (Lea, 1976d).

While several possible ways of using prosodics in parsing and the control strategy of a speech understanding system were suggested, time pressures in the ARPA program forced us to try only a limited form of prosodic aid to parsing. We had never before tested the idea of using prosodic information to directly aid parsing, so we confined our efforts in this initial study to the use of one simple prosodic feature, namely, intonationally-detected phrase boundaries.

Most readers may be aware that we have previously developed a computer program to detect boundaries between major syntactic constitutents, using fall-rise valleys in fundamental frequency. Figure 1 shows a typical example of how this program works, using one of BBN's travel management sentences. The program finds phrase boundaries at the valleys in fundamental frequency shown by the vertical lines, before the words "Lynn's" and "ASA". This program has previously been found to correctly detect about 90% of the major syntactic boundaries in connected speech. Recently (Lea, 1975b), we established that the <u>location</u> of the detected boundary is just before the first stress in the following constituent, regardless of whether that stress is in phrase-initial position or not. One problem in using intonationally-detected phrase boundaries is that this position may not line up with the exact <u>location</u> of the boundary between phrases as predicted by syntax. We were able to get around this problem by associating the boundary not with the first word in a phrase, but rather with the first <u>stress</u> in the phrase. To see how this works, we must first consider

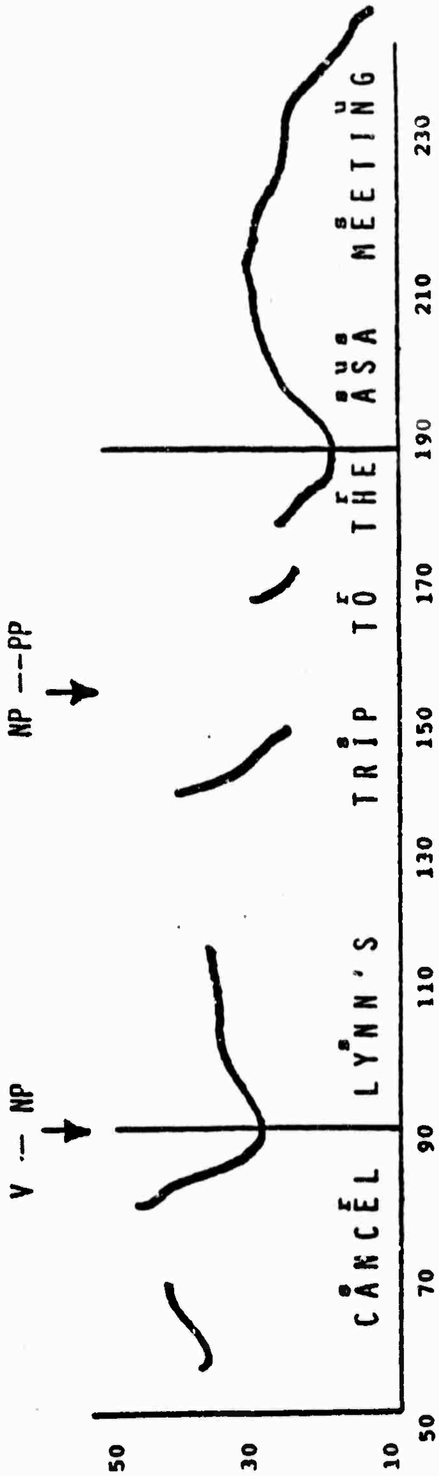

Figure 1. Phrase Boundaries (Vertical Bars) are Detected at Fall-Rise Fo Valleys in this Schematic of the Fo Contour of a BBN Travel Management Sentence

how intonationally-detected phrase boundaries can be introduced into the overall operation of the HWIM system.

Figure 2 shows a simplified diagram of the relevant aspects of the HWIM system. The system does a phonetic analysis to hypothesize words that have a high score of correspondence with the acoustic data. A queue of high scoring words, and their positions, is given by way of the system control to the parser. These one-word "islands" or theories can then be evaluated by the parser, and an adjusted list of likely words transferred back to system control. Prosodics may be introduced into the process, by adjusting the priority order of promising words that the parser transfers to system control.

In our initial experiment last May, fifteen sentences were processed through Sperry Univac's implementations of a pitch tracker and procedure for obtaining phrase boundaries. These prosodically obtained boundaries were then to be compared with expected boundaries, to see how well the prosodic structure of an utterance agreed with that expected for hypothesized word sequences. To exactly specify how expected boundaries relate to various syntactic structures, we marked those arcs in the SMALLGRAM grammar whose words are expected to be immediately preceded by Fo-detected phrase boundaries. Figure 3 displays a sample portion of the SMALLGRAM grammar, showing by the crosshatched lines those arcs that are expected to be accompanied by phrase boundaries. This noun phrase constituent provides descriptions of meetings which one might travel to. Its stressed beginnings are expected to be accompanied by a boundary. Thus, to describe the phrase "the last ASA meeting", we take the arc labeled "the", then the arc labelled "last", which is expected to be accompanied by an Fo-detected phrase boundary. Then the "sponsor" arc is taken, with the particular sponsor being "ASA", and finally the arc labelled "meeting" completes the phrase, and the grammar "pops up" to handle larger phrases. For the phrase "the ASA meeting", the grammar takes the "the" arc, the jump arc, then the sponsor arc for "ASA", which should be accompanied by a boundary, and again the "meeting" arc. Other stressed beginnings of the noun phrase shown at the bottom of the figure are also expected to be marked by boundaries. These include a sponsor like "ARPA" in a sentence like "John's going to ARPA", or a proper noun for a destination like "Sperry Univac".

In general, those arcs of SMALLGRAM which were marked by expected boundaries included ones labelled with main verbs, the first stressed words in noun phrases

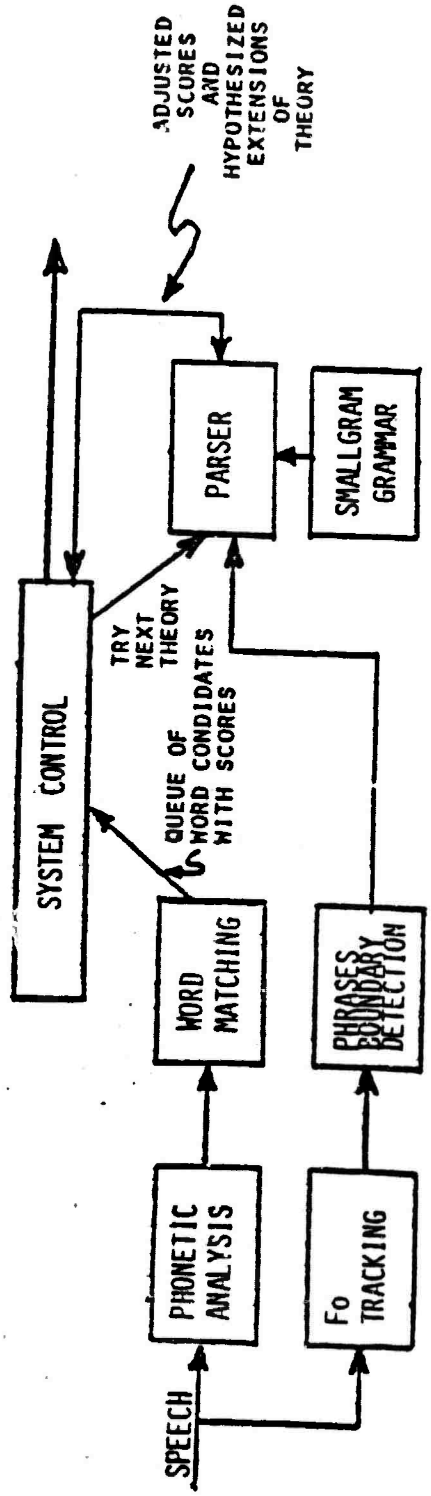

Figure 2. Simplified Schematic of How Phrase Boundaries are to Guide the HWIM Parser.

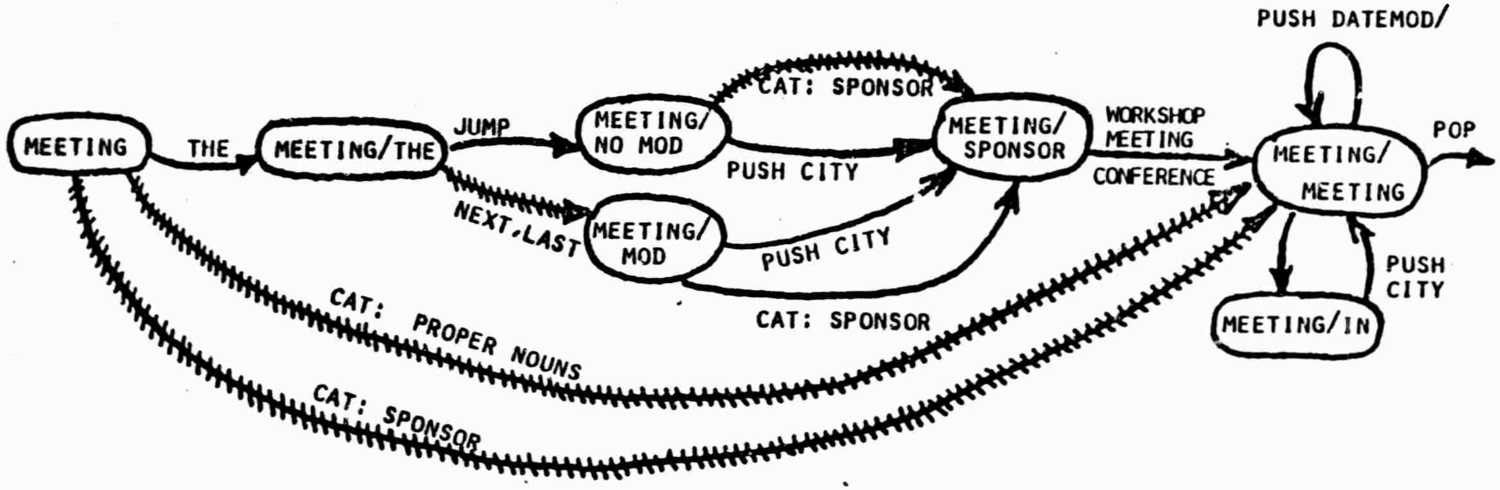

Figure 3. A Sample Portion of the SMALLGRAM Grammar Showing by Crosshatching Those Arcs that are Expected to be Accompanied by Intonational Boundaries

(i.e., quantifiers, adjectives, and nouns, when in first-stress positions); adverbs, and some special categories (like city, month, duration, digit, first name, etc.) that are used in BBN's travel-management task. My previous experiments had shown that Fo-detected boundaries would occur immediately before the first stressed syllable of such constituents.

Only about one out of every six arcs in the grammar is expected to be preceded by an Fo-detected boundary (Lea, 1976c, p. 10), yet at least once or twice in almost every sentence a boundary should occur, so that prosodic information should be helpful in almost every sentence.

Given that arcs in the grammar are marked if they are expected to be preceded by an Fo-boundary, we then need to determine whether the detected prosodic pattern for a sentence does or does not agree with the expectations about boundary placements, and to use that prosodic evaluation as a guide as to whether or not certain arcs should be used at an early stage in parsing. For our initial test last May BBN supplied complete control and parsing traces for seven of the fifteen sample sentences recorded for their travel-management task. We processed all fifteen sentences through our fundamental frequency tracker and obtained Fo-detected phrase boundaries, but could relate those results to syntax in only the seven sentences for which control and parsing traces were available. If an arc was marked to be preceded by a boundary, and a boundary was detected, then the priority, or "score", for the word on that arc was increased substantially (+50 points) whereas an expected boundary that was not detected caused the score of the word on that arc to be substantially decreased (-50 points). We thus reward theories of hypothesized words when those theories should be and are preceded by phrase boundaries, and we decrease the priority of words or theories that predict the presence of boundaries that are not detected. Within the BBN System, the adjusted score determines the priority of trying that theory in the subsequent analysis, with highest-score theories tried first.

Figure 4 displays the major reordering which prosodic boundaries can produce, for initial one word theories provided to syntax by the system control. For this one sentence ("Charge Bonnie's trip to budget item five."), the initial list of priorities or scores for one-word theories as provided by the original system is shown on the left, while its reordered form after prosodic adjustments is on the right. Notice that several instances of the correct words are moved up the list, so that four of the top six theories are now versions

INITIAL LIST	REVISED LIST
1. -PAUSE- 0 1 212	1. -PAUSE- 0 1 212
2. CHARGE-ED 0 5 197	2. CHARGE--ED 0 5 197
3. CHARGE 0 5 155	3. ITEM 22 26 193
4. ITEM 22 26 143	4. BONNIE 6 10 163
5. CHARG-ED 0 5 134	5. ITEM 23 26 158
6. CAR 0 4 129	6. CHARGE 0 5 155
7. ARE 24 128	7. CHARGE--ED 0 5 134
8. OUR 24 128	8. CAR 0 4 129
9. CHARGE-ED 1 5 125	9. ARE 2 4 128
10. HER 0 4 114	10. OUR 2 4 128
11. BONNIE 6 10 113	11. CHARGE-ED 15 125
12. THEIR 0 4 109	12. HER 0 4 114
13. THERE 0 4 109	13. THEIR 0 4 109
14. ITEM 23 26 108	14. THERE 0 4 109
15. WILL 24 26 105	15. WILL 24 26 205

Figure 4. Prosodic Rearranging of the Queue of Syntactic Events

of correct words. The words are followed by two numbers describing their beginning and ending positions, followed by the word score. For the words that moved up the list the scores increased 50 points due to prosodic agreement with expectations. The reordering, which is typical of the kinds of contributions prosodics can supply, permits correct theories (such as the words "charge", "item", and "Bonnie") to be tried at earlier stages in the analysis. Testing of erroneous words like "car, our, her, their, etc." will be delayed by the prosodic rearrangement of theories, so that false (misleading) paths can be avoided.

Also, at later stages, other good multiple-word theories "or phrases" are boosted by knowing boundary placements. For example, at a later stage in the analysis, the word sequence "Charge Bonnie" was boosted because of the correct occurrence of an expected boundary before "Bonnie". Likewise, "item five" and "budget item" were boosted, by prosodic agreements.

Analysis of all seven control traces supplied by BBN showed many instances where boundary placements could reorder the selection of theories, and thereby, the prosodics could direct analysis toward correct islands first, potentially saving computations on lower-score erroneous theories, and more directly reaching the goal of a correct parse. Without an index to the grammar (to know everywhere in the grammar each word could occur) and a complete list of theories (not just the top 15 such as were given on the available traces at each point in a parse), we could not be completely sure of when prosodies might make a parse succeed where it had previously failed. Also, without a full understanding of the timing of various processes, we couldn't tell exactly what amount or proportion of time could be saved by the use of boundary locations. Yet, it seemed clear that many time-consuming tests of erroneous paths would be eliminated or delayed until after better paths were analyzed when prosodies were used. The final test would come when the proposed procedures for prosodic aids were implemented and tested in a total system. Parsing and control traces with the use of prosodics could be compared with traces when prosodic information is not used.

2.2 Further Work Completed During This Reporting Period

Our analyses of how prosodics would rearrange the theories in available control and parsing traces clearly confirmed the value of incorporating prosodic

information into speech understanding systems. We then began a renewed effort to cooperate with BBN in incorporating these ideas into their system, for actual tests of whether parsing is more successful with prosodic information than without it. Our hope was that, before the 1976 version of the BBN HWIM system was to be demonstrated in September, we would have implemented and tested within that system all the procedures needed to use prosodic boundaries to aid the parser.

We received from BBN an updated version of the SMALLGRAM grammar, a grammar index to establish everywhere each word appeared in the grammar, and control and parsing traces for ten additional sentences. We marked _every_ arc in SMALLGRAM, showing one of four degrees of confidence that boundaries would occur with each of those arcs. We checked and refined these predictions, based on the additional evidence of detected boundaries in our prosodic processing of nine of the ten additional sentences. With BBN, we decided that quality scores on theories would be changed ± 50 points for those words that had the highest confidence level, and ± 20 points for the next highest confidence level. These changes in quality scores on words would then be reflected in subsequent calculations of shortfall densities within their new scoring strategy. We provided to BBN a specific assessment of how our prosodic programs had changed since they received a version over a year ago, and recommended minor changes in their versions to bring them up to date.

We recommended to BBN some necessary changes in the grammar. For example, some arcs previously had stressed words like "costs" on the same arc as unstressed words ("is", "was"), but, since boundaries would precede only the stressed words, the words should be separated onto two parallel arcs. Other changes in the grammar were needed to handle context-sensitive cases where a boundary occurred before a word if that word was not preceded by another (stressed) word that would pull the boundary earlier. Thus, for example, if an utterance includes the word sequence "the current budget", then a boundary occurs only before "current" (with none before "budget"), but if the utterance has the "the budget", then the boundary does occur before "budget".

In July, we supplied all the new prosodic results, the completely marked grammar, the updated scoring adjustment procedures, and the grammar changes, to BBN. We also supplied copies of portions of the traces for the nine sentences,

showing the reordering that would occur if ± 50 point adjustments in quality scores were made where boundaries were expected and found or not found. The initial one-word scans or queues were significantly improved by the introduction of boundary information, with more correct words moved to high priority positions in the lists, and many incorrect words reduced in priority. A few erroneous words were increased in priority.

Several subsequent queues of events with multiple-word theories were also provided, showing increased priority for correct theories, although the exact reordering for multiple-word theories would require full calculations of MAXSEGS, shortfall, and shortfall density values used in the new BBN scoring procedures. These later queues from parsing traces also graphically illustrated how elimination of one wrong theory from the early stages of analysis could eliminate many wrong theories in later queues.

One other critical task was to precisely define the time limits for location of boundaries that may be said to "immediately precede" a marked arc of the grammar. Since our previous research had shown that, with rare exceptions, the Fo-detected boundary was located immediately before the first stressed syllable in a phrase, BBN (specifically Jerry Wolf), with Sperry Univac's advise, developed the following procedure for relating syllable boundaries, word boundaries, and detected phrase boundaries. First, each detected phrase boundary is "smeared" over the two closest syllable nuclei, by bracketing the boundary position between the beginning of that syllable whose nucleus center is immediately before the detected time and the ending of that next syllable, whose nucleus center just follows the boundary time. This thus requires that the system not only use the Fo-based boundary detection, but also the syllabification routine ("CHUNK"). Next, (with one exception soon to be described) the word being hypothesized by the system is said to be immediately preceded by that Fo-detected phrase boundary if the time of the beginning of the word lies within the phrase-boundary-smeared region (between the one syllable onset and the end of that next syllable). If the word is not stressed on its first syllable in the lexicon, then a match of boundary location and the word location is considered to be acceptable as long as there is any overlap of the time the word spans and the time the smeared boundary spans.

All of the above ideas were implemented within the BBN system, but time pressures at the end of BBN's ARPA contract unfortunately did not permit testing out these prosodic aids to parsing. The final test would have come if parsing and control traces _with_ the use of prosodics could be compared with traces when prosodic information is _not_ used. The project terminated before any such results were available. Still it seems likely from our hand analyses of sixteen BBN control traces that prosodically-detected phrase boundaries can improve the order in which alternative syntactic theories are tested. While we obviously have just begun to offer specific prosodic aids to parsing and system control, I believe we have shown enough of a pattern of improved ordering of alternative theories and events to warrant future efforts to introduce prosodics into parsing procedures in speech understanding systems.

3. ACOUSTIC-PROSODIC PATTERNS IN 255 SENTENCES

In previous reports (Lea, 1973a, 1973b, 1974, 1975a, 1976b; Lea, Medress, and Skinner, 1973a,b; Lea and Kloker, 1975), we have reported on a variety of experiments that demonstrate the utility of prosodic information in speech understanding and have indicated various prosodic regularities such as the marking of phrase boundaries by Fo valleys just before the first stress of a phrase, and the marking of stressed syllables by a complex combination of pitch rises and high energy integrals. We argued that our initial studies with uncontrolled speech texts needed to be supplemented by much more controlled studies with sentences which carefully isolated various linguistic contrasts and their prosodic correlates. A list of 1100 sentences, each spoken by three talkers, was designed and recorded to provide the necessary data for more controlled studies.

3.1 Previous Determination of Sentences and Perceived Stress Patterns

We began our studies of prosodic regularities in the database sentences by selecting 255 sentences, spoken by one talker, as a representative sampling of linguistic contrasts that were to be studied (Lea, 1976c). These sentences, listed again in Appendix A, include ones which can test the prosodic effects of moving the first stress in a constituent from the first, to the second, to the third (etc.) syllable (to see how the Fo-detected phrase boundary was then positioned); expansions of noun phrases to include nouns, pronouns, articles, quantifiers, adjectives, and participles; practically-identical word sequences with contrasting syntactic structures, verb-versus-noun stress pairs (permit/permit, etc.); NP-PP-PP subordination; various sentence types (declarative, WH question, yes/no question, command); coordinate clauses, VP's, and NP's'; and relative clauses.

Presented in our last report (Lea, 1976c) were listeners' perceptions of stress patterns in the 255 sentences. We confirmed the consistency of listeners' perceptions from time to time and from listener to listener, and showed which word categories are usually stressed, which are perceived as unstressed, and which are reduced. We showed how coordinate structures and subordination produce reduced stress on verbs, function words, and repeated nouns. Perceived stresses were found to consistently decrease throughout a word sequence without right syntactic brackets, even though nuclear stress rules predict patterns of increasing stress.

In the remainder of section 3, we will present results about the acoustic-prosodic patterns in those same 255 sentences. We will see how various prosodic hypotheses stand up to the data from these carefully controlled structural distinctions, and what are the acoustic correlates of the perceived stress patterns. Some improvements in the prosodic analysis tools were needed for these studies, and they are briefly described in section 3.2. Then, in section 3.3, we will discuss the regularities in placement of Fo valleys associated with phrase boundaries, and present the level of performance of our computer programs for detecting phrase boundaries. These designed sentences help firmly answer specific questions about which constituents are marked by Fo-detected boundaries, and where the boundaries occur (cf. Lea, 1975b, for related work on some others of the database sentences). Section 3.4 covers performance of the stressed syllable location algorithm, and the prerequisite process of syllabification. Some intonational hypotheses that could be tested with the data are discussed in section 3.5. More detailed studies of acoustic correlates of stress in various sentence structures are suggested in section 3.6. Specific implications for using these experimentally-verified prosodic regularities in speech understanding systems are discussed in section 3.7.

3.2 Improved Methods for Acoustic-Prosodic Processing

In earlier studies (Lea, 1973f), we used a software analysis to obtain the sonorant energy function from the values of LPC-analyzed spectral energy in the band 60 Hz to 3000 Hz. Last year, we introduced an analog hardware filter to replace those spectral computations (Lea, 1975b), but because of the noisy signal and other minor problems obtained with the filter, plus recent improvements in the speed of computation of our spectral analyses, we have returned to the use of the software-derived sonorant energy function.

Our Fo tracker, which is based on an autocorrelation analysis (Skinner, 1973), has been improved by incorporating a weighting function with a 10% slope on the autocorrelation function, to discourage octave errors. Thus, the autocorrelation equation was given by

$$A_i = \sum_{L=1}^{N} C_i \cdot C_{i+j-1}$$

from $j = O_L$ to $j = O_M$, where i is the amount of offset, N is the length of the window, C_i is the windowed and center clipped waveform, C_{i+j-1} is the offset window to be correlated with C_i, and O_L and O_M are the lower and upper autocorrelation offset limits, respectively (See Lea, Medress, and Skinner, 1973a, p. 31).

We next compute a straight line approximatation to a logarithmic function that decreases the A_i's by WGT percent per octave:

$$SLOPE = \frac{-(\text{Log}_2 O_M - \log_2 O_L)}{O_L - O_m} \cdot \frac{WTG}{100}$$

We are currently using WTG = 10. Figure 5 shows a sketch of the weighting factor $(1 + (i-O_2 + 1)\ (SLOPE))$, which is based on this logarithmic weighting function.

Finally, we weight the autocorrelation coefficient by that straight line:

$$B_i = A_i\ (1 + (i - O_L + 1)\ .\ SLOPE)$$

for each i, $i = O_L$ to O_M. This causes the kind of change in the autocorrelation function sketched in Figure 5, in which the larger offset is selected as the peak in the autocorrelation function after the weighting, whereas the smaller offset would have been erroneously selected before the weighting was applied.

Analysis of the 255 sentences showed that some octave errors still occasionally occur, but much less frequently than without this weighting function.

The program for detecting phrase boundaries from Fo valleys (Lea, 1975), is unchanged except for using a threshold value of 4 rather than 5 eighth tones change in Fo to have a substantial Fo valley.

Recently, we changed our syllabification procedure in several minor ways and one major way. The minor changes included: requiring all frames within a syllabic nucleus to be voiced, disallowing overlapping nuclei that occasionally occurred; setting a minimum energy level for a peak in energy to qualify as the peak of a syllable; smoothing one-point jumps or dips in sonorant energy that occasionally caused erroneous syllable detections; and adjusting the threshold value for a minimum dip and rise in sonorant energy to qualify as a syllable boundary (now 4 dB rather than 5 dB), to help pick up some missed syllables. These changes helped reduce the number of erroneous "syllables" detected and helped locate some syllables that had previously been missed.

The major change in the syllabication procedure was to redefine the beginning and ending of a syllabic nucleus. Previously the nucleus endpoints were defined as those points where energy first dipped 4 or 5 dB below the peak energy level in the syllable. This often excluded from the syllable nucleus

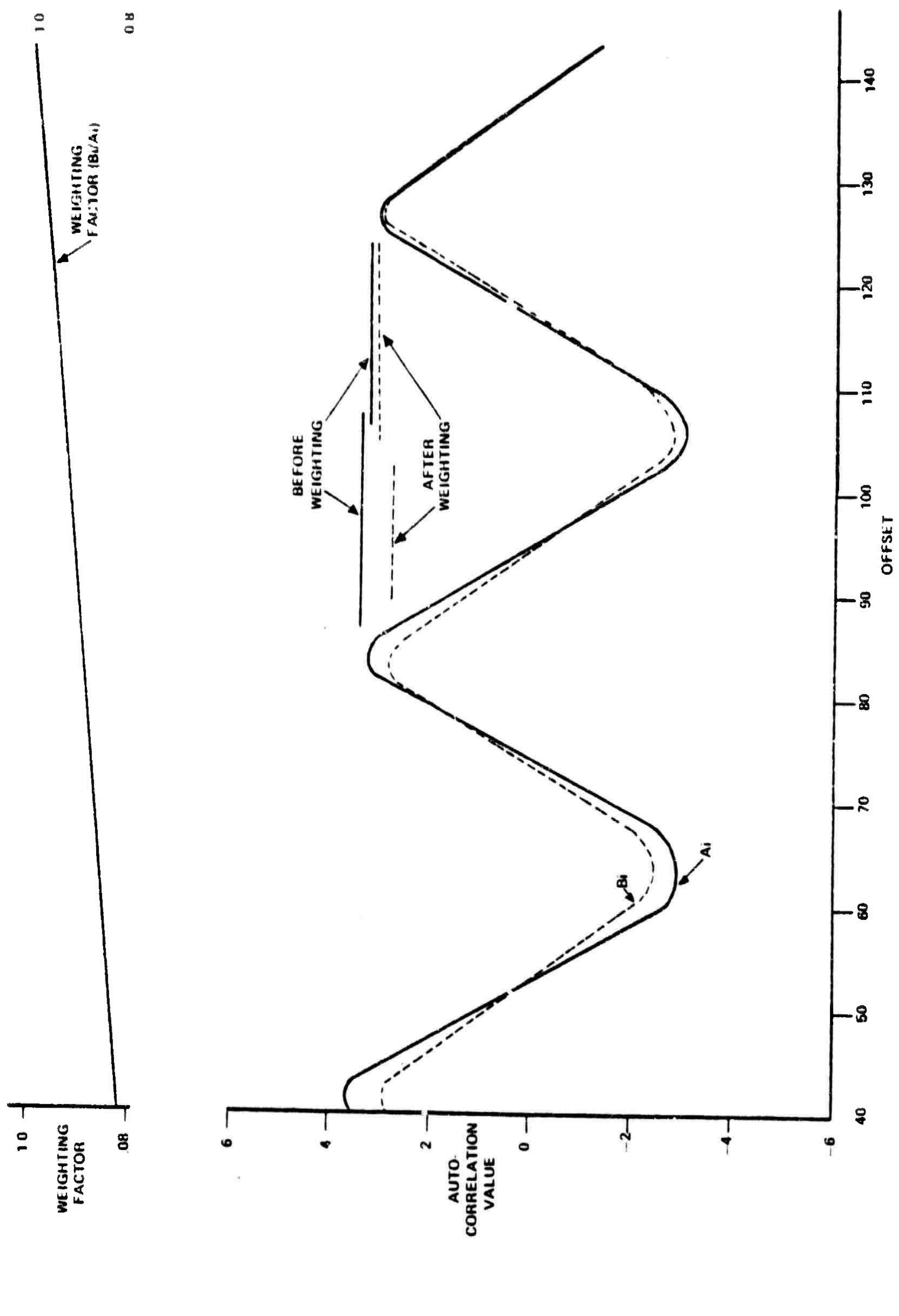

Figure 5. Use of a Weighting Factor to Change the Autocorrelation Values and Avoid Octave Errors

much of the non-vowel sonorant portions of the syllable, and even some of the lower-energy regions of the vowel. An investigation showed that the syllable nucleus limits lined up best with the endpoints of sonorant phonetic segments in available transcriptions when the syllable nucleus limits were defined as: those points, as one moves either way from an energy dip at a syllable boundary, where energy first rises to at least one half of the distance between the low energy in the dip and the peak energy in the syllable nuclei. That is, if D(i-1) is the energy level in the dip before a syllable nucleus, D(i) is the energy level in the dip after the nucleus, and P(i) is the peak energy level within the nucleus, then the beginning of the nucleus is now the first point where energy E satisfies the following inequality:

$$E > (P(i) + D(i-1))/2.$$

The endpoint of the nucleus is the last point in time before the following dip at which energy E statisfies the following similar inequality:

$$E > (P(i) + D(i))/2.$$

A major revision has been made in the nucleus finding subroutine to accomplish this change, and it has been tested and found to be working correctly. One obvious complication that this introduces is that the nucleus durations no longer agree with those previously expected by the stress identification subroutine, so that some errors in stress location may result. This requires some further testing and possible changes in the location routine, to correctly select stressed nuclei.

Besides these revisions in syllabification, we also made minor improvements in the stress finding routines, removing a few minor bugs and also modifying the procedure for defining the position of peak fundamental frequency (Fo) in a phrase. With this modification, local Fo jumps after unvoiced consonants do not cause a displacement in the peak-Fo point before which we search for the first stress in the phrase.

A preliminary test showed that the high-frequency (650 - 3000 Hz) sonorant energy function, when used in place of the regular (60 - 3000 Hz) sonorant energy function, could detect some syllable boundaries at intervocalic non-vowel sonorants, but it also was less smooth than the regular sonorant energy function, and introduced some erroneous syllable boundaries. Using both functions together, or using some other spectrally-weighted energy function, may

permit the detection of more syllabic boundaries without introducing erroneous boundaries.

3.3 Intonational Phrase Boundaries

The results in syntactic boundary detection from Fo contours are summarized in Table I. Each subset of sentences (listed completely in Appendix A) is listed separately, with a brief explanation of the structures it tests, the number of expected (syntactically-predicted) boundaries in the subset, and the numbers of correctly-detected boundaries, extra (syntactically-unexpected, but still apparently syntactically-related) boundaries, and false (phonetically-produced) boundaries. The overall figure of 76% of all expected boundaries being correctly detected compares fairly well with scores ranging from seventy percent to over ninety percent found in our previous studies (Lea, 1972, 1973c, 1976c). This is very encouraging, in light of the difficult cases for boundary detection included in this database. Many of the sentences are all-sonorant to avoid false boundaries associated with Fo changes near obstruents. However, this decreases the likelihood of sufficient Fo variations at expected boundary positions. Also, many sentences are very short (e.g., in subsets 1A, 1E, $2C_2$, 3F, 6A, 7B, and other scattered sentences), and previous experience (as well as published claims; Armstrong and Ward, 1929) predicts that such sentences will be more monotone and not as likely to exhibit the substantial Fo variations needed for boundary detection. We could lower the threshold for a substantial valley to 4 eighth tones Fo dip and rise in this study, precisely because the less frequent occurrence of obstruents meant less likelihood of false phonetically-produced boundaries.

One can use the boundary detection results for the individual sentences and subsets in this study to evaluate our basic boundary detection hypothesis and to more firmly establish exactly which constituents are accompanied by Fo-detected boundaries. Our basic boundary detection hypothesis can be stated as follows:

> A substantial (4 eighth-tone) Fo valley will occur just before the first stressed syllable of each major syntactic constituent which is preceded by another constituent containing a stress. Major syntactic constituents include: main verbs; auxiliary verbs <u>if and only if</u> they contain a stress, such as a negative; noun phrases; sentence adverbs; conjuncts; relative clauses; and parentheticals. In addition, a boundary will occur before a preposition if it is stressed (besides the one that will occur before the stressed noun phrase within the prepositional phrase). A boundary will occur between the parts of a compound noun.

TABLE I. SUMMARY OF BOUNDARY DETECTION RESULTS

	SUBSET, AND STRUCTURES STUDIED	BOUNDARIES EXPECTED	HITS		EXTRA		FALSE	
			N	%	N	%	N	%
1A	Stress Movement, 1 Stress/constituent	44	21	46%	-	-	5	19%
1B	Stress Movement, Expanding Determiner	26	20	77%	3	13%	-	-
1C	Stress Movement, NP Modifiers, <4 S/C	50	46	92%	6	11%	3	6%
1D	Stress Movement, NP Modifiers, >4 S/C	54	49	91%	25	32%	3	4%
1E	"Flying Planes" Paradigm	12	12	100%	--	--	--	--
$2C_2$	Stress Movement in First Constituent	30	19	63%	2	9%	3	13%
3D	Verb/Noun Stress Pairs	58	56	97%	15	18%	24	28%
3F	Phonetic Influences	28	18	64%	3	13%	3	13%
4C	NP-PP-PP Subordination	25	19	76%	2	9%	1	5%
6A	Commands	33	27	82%	1	4%	--	--
7B	Yes/No Questions	18	10	56%	14	54%	2	8%
7D	WH Questions	46	33	72%	2	6%	--	--
8A	Coordinate Sentences	111	71	64%	4	5%	--	2%
8H	Coordinate Verb Phrases	55	46	84%	4	8%	1	2%
8K	Coordinate Noun Phrases	36	33	82%	9	20%	3	7%
11A	Relative Clauses	60	39	65%	10	18%	8	14%
	TOTALS FOR ALL SUBSETS	686	519		100		57	
	OVERALL PERCENTAGES			76%		15%		8%

Let us consider what each of the subsets teaches us about the adequacies and inadequacies of this hypothesis. Subset 1A, for example, has very simple short sentences of the forms (NP V), (NP V NP), and (NP AUX V NP). We expect a boundary before the main verb (V) and before the second NP. In fact, only 8 of the 24 main verbs in subset 1A were preceded by boundaries. Five of the eight were before the word "worry", in fact all of the cases where the verb was "worry" were preceded by boundaries, while the verbs "know ", "knew", "owe", and "enroll" were not preceded by boundaries, unless there was an (unstressed) auxiliary preceding those verbs. Is there something about the phonetic structure (prestressed /w/ versus other sonorants) that causes this apparent lexical difference? Or, is the difference due to the different stress patterns of the verbs (SU versus either S or US)? Certainly the presence of boundaries before a main verb when the verb is preceded by an unstressed auxiliary suggests that the alternating stress pattern SUS is more likely to have a boundary before the second S than the pattern SS. This is to be expected, since Fo is usually lower in unstressed syllables than in stressed ones, so the Fo valley should more readily occur when an unstressed syllable intervenes between stresses. This explains the three cases where a main verb (other than "worry") preceded by an auxiliary was accompanied by a boundary. Still, we would expect, according to the basic boundary-detection hypothesis, that a boundary should occur in the SS sequence, between the stressed noun subject of the sentence and stressed verb. We have previously tried (Lea, 1973b) to accept the counter-hypothesis that no boundary will occur between a noun and a following verb, but the five cases of boundaries with "worry" discount that counterhypothesis.

Subset 1A also showed that 12 of the 18 expected boundaries between main verbs and following object NP's were detected. All those that were missed had SS sequences at the verb-noun boundary, suggesting again that boundaries are less reliably detected within SS sequences. However, there were four SS sequences at the V-NP boundary that were accompanied by detected boundaries. We can say that if the basic boundary detection hypothesis predicts a boundary and the phrase boundary is spanned by an SUS sequence, the boundary will be detected, if the boundary is immediately preceded and followed by stresses it may or may not be detected.

These uncertainties suggest that more data needs to be examined, to determine what causes some but not all NP-V and V-NP boundaries to be detected. Table II shows the results of analyzing all NP-V and V-NP boundaries in the 255 sentences, excluding those cases where other factors like coordinate NP's, relatives in the NP, or other structural issues might interfere. (Note

TABLE II. CORRECT DETECTIONS OF NP-V AND V-NP BOUNDARIES IN THE 255 SENTENCES

SUBSET NUMBER, AND STRUCTURES STUDIED		NP-V BOUNDARIES		V-NP BOUNDARIES	
		SS SEQUENCES	SUS SEQUENCES	SS SEQUENCES	SUS SEQUENCES
1A	Stress Movement, 1S/C	4/15=27%	4/9=44%	3/9=33%	8/8=100%
1B	Stress Movement, Expanding Determiner	---	10/33=77%	9/11=82%	1/2=50%
1C	Stress Movement NP Modifiers, <4S/C	---	22/25=88%	13/14=93%	11/11=100%
1D	Stress Movement, NP Modifiers, >4S/C	---	23/27=85%	19/20=95%	7/7=100%
1E	"Flying-Planes" Paradigm	---	4/4=100%	---	8/8=100%
$2C_2$	Stress Movement in 1st Const.	1/4=25%	9/11=82%	2/6=33%	7/9=78%
3D	Verb/Noun Stress Pairs	---	13/14=93%	2/2=100%	9/9=100%
3F	Phonetic Influences	2/5=40%	3/4=75%	10/12=83%	---
4C	NP-PP-PP Subordination	---	2/5=40%	5/5=100%	0/2=0%
6A	Commands	---	---	5/8=63%	7/7=100%
7B	Yes/No Questions	5/5=100%	1/5=20%	3/6=50%	1/1=100%
7D	WH Questions	4/8=50%	3/7=42%	4/7=57%	4/4=100%
8A	Coordinate Sentences	1/16=6%	14 24=58%	22/32=69%	6/8=75%
	TOTALS FOR ABOVE SUBSETS	17/53=32%	108/148=73%	97/122=79%	69/76=91%

that SU*S means cases where one or more unstresses occur between the last stress of the previous phrase and the first stress of the next phrase.) The totals for all subsets, and the results for the individual subsets, show that both the NP-V and V-NP boundaries are less likely to be evident when two stresses are adjacent, while intervening unstressed syllables aid the Fo-marking of boundaries. Yet, this effect is much more pronounced for NP-V boundaries than for V-NP boundaries. In this sense, the V-NP boundary is a much more stable, reliably detected boundary than is the NP-V boundary. Contrary to the assumption in many published works (Lieberman, 1967, Scholes, 1969; Oller, 1973;), the subject-predicate (or NP-V) boundary is not the most robust or prominent boundary in acoustic data (cf. Lea, 1972, 1973a).

We may very well be disappointed that the acoustic detection of phrase boundaries cannot be simply explained in purely syntactic terms, without the need for disclaimers or qualifying phrases related to lexical choices, stress patterns, or phonetic structures. But, until we have an adequate model of all influences on Fo contours, we apparently must acknowledge such loose generalities as the notion that boundaries are "more likely to be detected" when unstresses intervene between the last stress of one phrase and the first stress of the next phrase.

There are a number of other interesting results embedded in the figures of Table II. For example, the rightmost column of the table shows that excellent performance in boundary detection was obtained with V-NP boundaries accompanied by SU*S sequences. Two of the seven missed V-NP boundaries with SU*S sequences were accompanied by borderline Fo variations of 3 eighth tones. Four others involved words (repeated nouns and the command verb "put") that were predicted to be stressed, but the listeners did not actually hear as stressed, so no boundary should be expected. Thus, there are actually only 3 out of 72 expected V-NP boundaries with SU*S sequences that were not detected. Some of the other lower scores in the table result from cases where predicted stresses were not actually perceived as stressed. For example, lower scores with coordinate sentences (subset 8A) resulted from predicting that the repeated verbs and nouns would be stressed, when in fact repeated words were usually not stressed. The possessive pronoun "mine" was predicted to be stressed for the commands in subset 6A, but several boundaries were not detected before those pronouns since "mine" was not actually stressed as expected. A glance at the stress patterns and boundary

results shown in Appendix A for each individual sentence can assure one that a significant fraction of the "missing" boundaries <u>should</u> have been missing since the surrounding words were not stressed as expected.

Boundary detections for other types of phrase structures are given in Table III. Clearly, most other types of expected boundaries are very reliably detected. The only V-PP boundary missed was in a relative clause, which had reduced stresses and a fairly flat Fo contour. Over half of the AUX-V boundaries missed were in coordinate structures where the verb was not perceived as stressed, and thus these structures would <u>not</u> be expected to be accompanied by boundaries. All of the missing NP-PP boundaries were before a short utterance-final PP ("from Maine", "of May") with weak stress on the noun, and consequently, fairly flat Fo contours. The two missing NP-NP boundaries also involved low stresses due to occurrence in a coordinate structure or a relative clause. The two missing NP-ADV boundaries were before moderate utterance-final stresses, but exhibited 3-eighth-tone Fo valleys (almost sufficient to be detected as boundaries). Both of the cases where a relative pronoun was not followed by a boundary had an unexpected boundary <u>before</u> the relative, and small Fo valleys after the relative pronoun. Two boundaries before stressed auxiliaries (involving negatives) also were not found, apparently due to reduced stresses in coordinate constructions. Clearly, one of the primary causes for missing some expected boundaries is the reduced stresses that accompany subordination, relative clauses, and repetition in coordinate constructions. Comprehensive stress rules could predict the absences of stresses in such structures, and no boundaries would be expected in such circumstances.

Table IV summarizes where extra, unexpected boundaries occurred in the 255 sentences. Several regularities suggest the value of modifying the current phrase boundary location hypothesis (page 26). In particular, there seems to be a regular occurrence of a boundary before the noun (or, if the noun is an unstressed pronoun, before the verb) in a yes/no question, even though the initial auxiliary verb is perceived as unstressed or reduced. Perhaps such boundaries should be expected Similarly, while our basic hypothesis that there is no boundary internal to an NP (except for compounds) seems to hold in the majority of cases, the fact that two thirds of the ADV-ADJ boundaries are accompanied by Fo valleys suggests that perhaps those boundaries should be expected. Finally, we might revise the hypothesis to predict a detected boundary between a NP and its following relative pronoun. This seems particularly strange, since the relative pronoun is heard as unstressed.

TABLE III. CORRECT DETECTIONS OF VARIOUS TYPES OF EXPECTED BOUNDARIES

PHRASE STRUCTURE TYPE	BOUNDARIES EXPECTED	HITS	PERCENTAGE CORRECTLY DETECTED
Boundaries At Verbals:			
V-ADV	2	2	100%
V-PP	10	9	90%
AUX-V	120	96	80%
Boundaries At NP's:			
N-N Compounds	7	7	100%
Preposed ADV-NP	2	2	100%
NP-PP	32	28	88%
NP-NP	17	15	88%
NP-ADV	13	11	85%
Relative Pronoun-NP	10	8	80%
NP-(AUX+NEG)	7	5	71%
Conjoined Structures:			
(VP Conj)-VP	12	12	100%
NP,-(NP,Conj NP)	5	5	100%
VP,-(VP, Conj VP)	3	3	100%
(S Conj) - S	22	21	95%
(NP Conj) - NP	18	17	94%

TABLE IV. UNEXPECTED (EXTRA) BOUNDARY DETECTIONS

PHRASE STRUCTURE TYPE	OCCURRENCES OF THE STRUCTURE	NUMBER OF BOUNDARIES DETECTED	PERCENTAGE OF OCCURRENCES THAT WERE DETECTED
Boundaries At Verbals:			
(AUX+PRONOUN)-V in Y/N?	3	3	100%
AUX-NP in Y/N?	8	6	75%
NP-AUX	125	4	3%
Boundaries Within NP's:			
ADV-ADJ	12	8	67%
ADJ-N	69	22	32%
QUANT-ADJ	19	6	32%
QUANT-N	12	3	25%
ADJ-ADJ	27	6	22%
Other Boundaries:			
NP-Relative Pronoun	17	15	88%
Within Last Word of Y/N?	14	5	36%

Interestingly, eighteen (i.e., 40%) of the unexpected boundary detections in NP's involve the words "moral" or "immoral", even though those words occur in only 20 (22%) of the multiple-stress NP's. Over 80% of the unexpected boundaries within NP's involved multisyllabic words (and thus alternating stressed-unstressed patterns), suggesting (as we found with the correctly detected NP-V and V-NP boundaries on pages 21 to 23) that unstresses between stresses increase the chance of boundaries being detected.

Also listed in Table IV are five cases where extraneous boundaries were found within the last word of a yes/no question. This was due to an Fo valley appearing just before the terminal rise in Fo that often accompanies yes/no questions.

One other category of boundaries that needs mention are the "false" boundaries listed in the rightmost column of Table I (page 20). Twenty false boundaries (35%) were in the initial syllable of an utterance, resulting from local Fo variations at voice onset (perhaps due to glottal stops or such). These could be eliminated by setting a minimum time between the onset of voicing in an utterance and the time of the first possible phrase boundary. Another thirty (56%) of the false boundary detections resulted from Fo variations associated with non-initial obstruent consonants. Other false boundary detections resulted from errors in the Fo contour, and phrase-final terminal rises in Fo. A graphic illustration of how false boundaries are introduced by presence of obstruents is that twenty four (42%) of the false boundaries were in subset 3D, which has many obstruents in it.

In summary, it appears we are very near optimal attainable performance in phrase boundary detection from Fo contours, with few possibilities of improvement by revisions in the computer program. For some ideas about further detailed improvements in the BOUND 3 phrase boundary detection program, see our previous semiannual report (Lea, 1976c). However, results with the 255 sentences suggest that the places where boundaries should be predicted could deserve further study. We have found that expected stress patterns should be taken into account, so that boundaries will not be expected in coordinate structures with repeated words, or in some subordinate structures. Even the presence or absence of unstresses between stressed syllables could be used to refine the probability of detecting a phrase boundary. On the other hand, boundaries should perhaps be expected at the first stress after the auxiliary verb in a yes/no question, between a noun phrase and the relative pronoun of its subordinate relative clause, and perhaps even between an adverb and the adjective it modifies within a noun phrase. The minimal contrasts

in structure within pairs of the database sentences have been useful in highlighting these refinements. Further work, with other talkers and more structures, should be done.

3.4 Syllabification and Automatic Location of Stress

Syllables are located in the spoken sentences by detecting substantial (4dB) dips in sonorant (60-3000 Hz) energy which occur in the consonantal region of syllable boundaries. The syllabic nucleus (vowel and some adjacent sonorant consonants) is centered around the local peak in sonorant energy, with beginning and ending points of the nucleus located at those points, closest to the preceding and following dips, whose energy is at least half of the distance from the value in the dip to the value at the syllabic peak.

The 255 sentences provide very demanding tests of the syllabification procedure, since many sentences are all-sonorant. Sonorant consonants do not produce large energy dips such as obstruents do. A large majority of the failures to locate syllables in previous studies have been due to intervocalie sonorants not providing the necessary dips in energy, so two or more syllables in an all-sonorant sequence appear as one nucleus. (Most of these long combinations of syllables then appear as single stressed syllables.)

Table V shows the syllabification results for each subset of the 255 database sentences that have been studied. The overall result of 91% correct detection of expected syllables is very satisfying, particularly for this difficult data. Also, the predictions of expected syllables were biased against the syllabification procedure, in that words like "tower", "moral", and "eyeing" were predicted to be two syllables even though one could anticipate that the two syllables could merge into one in actual speech.

Fifty one (29%) of the 175 syllables that were not automatically detected were the weak syllables in five words (moral, immoral, Mary, marry, ruin) that were always detected as having less syllables than expected. Other less frequent words like "worry", "Armenian", "tower", "Murray", "Marion", "aluminum", "erring", etc., were also consistently found with fewer syllables than expected, and accounted for over 25 (14%) of the missing syllables. Other words that often, though not always, were found with fewer syllables than expected included "really" "marine", "airmen", "enroll", etc.

TABLE V. SYLLABIFICATION RESULTS

SUBSET NUMBER, AND STRUCTURES STUDIES		SYLLABLES EXPECTED	HITS N	HITS %	FALSE	FALSE %
1A	Stress Movement, 1S/C	102	87	85%	2	2%
1B	Stress Movement, Expanding Determiner	91	84	92%	2	2%
1C	Stress Movement NP Modifiers, 4S/C	192	176	92%	0	0%
1D	Stress Movement, NP modifiers, 4S/C	256	217	85%	1	0.5%
1E	"Flying-Planes" Paradigm	50	43	86%	0	0%
$2C_2$	Stress Movement in 1st Const.	106	83	78%	1	1%
3D	Verb/Noun Stress Pairs	236	222	94%	4	2%
3F	Phonetic Influences	54	54	100%	3	5%
4C	NP-PP-PP Subordination	86	85	99%	1	1%
6A	Commands	99	86	87%	0	0%
7B	Yes/No Questions	73	68	93%	1	1%
7D	WH Questions	99	97	98%	0	0%
8A	Coordinate Sentences	189	181	96%	1	1%
8H	Coordinate Verb Phrases	115	100	87%	1	1%
8K	Coordinate Noun Phrases	116	110	95%	2	2%
11A	Relative Clauses	120	116	97%	4	3%
TOTALS FOR ALL SUBSETS		1984	1809		23	
OVERALL PERCENTAGES				91%		1%

Some persistent tendencies to lose syllables by grouping two syllables across word boundaries were also found. Twenty three (13%) of the missing syllables were a result of the first two syllables of "will enroll" appearing to be one syllabic; "know a(n)" gave five other misses, while "enroll a(n)" gave seven more.

Table V shows that all-sonorant subsets 1A, 1D, 1E, $2C_2$, 6A, and 8H are the only ones with syllabification scores under 90%, while subsets with obstruents, like 3D, 3F, and 4C give above-average success in syllable detection.

Also of interest, are the 23 false alarms in syllable location, shown in the rightmost column of Table V. Nine of these appear to be due to a bug in the program which is currently allowing short (20 or 30 ms) chunks of high energy to be called syllabic nuclei.[1] There is a test in the CHUNK program that should be throwing out all syllable candidates whose nuclei are less than 40ms in duration Related to these are four other cases where erroneous utterance-final or phrase-final syllables are inserted due to our smoothing of the energy function, which improperly brings some energy levels within noise up to the high energies needed for syllable detection. Five other low-energy erroneous syllable detections could be eliminated by setting a threshold of minimum energy below which no syllabic could be detected. Three cases occurred where it appears the talker actually said an extra syllable, like "nine-uh" for "nine".

While the syllabification results are very satisfying, there is room for some future improvements. One promising idea is to modify the spectral weighting of the energy function so it dips more reliably and substantially in non-vowel sonorants at syllable boundaries. In some preliminary studies with other data, we found that a high-frequency (650-3000Hz) sonorant energy function dipped more in intervocalic sonorants than did the regular sonorant energy function, but it was noisier or more prone to introduce false boundaries within syllables. The two together, or some spectrally-weighted "loudness function" that dips more in intervocalic sonorants, would seem to be needed.

Once syllables are located, we can then determine which ones are stressed. Our STRESS program associates stresses with high-energy syllabic nuclei near the Fo rise at the beginning of an Fo-detected phrase, and at local inflections in Fo at later points in the phrases. The results of applying this program to the 255 sentences are shown in Table VI. The overall score of 92% correct locations of stressed

1 As this report went to press, this bug had been corrected, so that only 14 false alarms remain.

TABLE VI. STRESS LOCATION RESULTS

SUBSET NUMBER, AND STRUCTURES STUDIES		STRESSES	HITS N	HITS %	FALSE N	FALSE %
1A	Stress Movement, 1S/C	60	55	92%	7	11%
1B	Stress Movement, Expanding Determiner	51	49	96%	8	14%
1C	Stress Movement NP Modifiers, 4S/C	112	97	87%	18	16%
1D	Stress Movement, NP modifiers, 4S/C	138	120	87%	25	17%
1E	"Flying-Planes" Paradigm	24	24	100%	9	27%
$2C_2$	Stress Movement in 1st Const.	44	42	95%	8	16%
3D	Verb/Noun Stress Pairs	86	78	91%	34	30%
3F	Phonetic Influences	44	42	95%	3	7%
4C	NP-PP-PP Subordiantion	37	30	81%	8	21%
6A	Commands	57	56	98%	8	13%
7B	Yes/No Questions	38	33	87%	17	34%
7D	WH Questions	64	61	95%	6	9%
8A	Coordinate Sentences	93	88	95%	34	28%
8H	Coordinate Verb Phrases	57	53	93%	15	22%
8K	Coordinate Noun Phrases	54	50	93%	15	23%
11A	Relative Clauses	74	69	93%	23	25%
TOTALS FOR ALL SUBSETS		1033	947		228	
OVERALL PERCENTAGES				92%		20%

syllables is surprisingly good, and is comparable to our best results on previous studies with read speech. Equally satisfying is the fairly low percentage of stress locations that were false (i.e., located syllables that were not perceived of five listeners). In past studies (Lea, 1974a, p. 19), we have found comparable figures for such false alarms, but the new definition of nucleus durations extending out to the half-way-down points, with resulting longer nucleus durations, was expected to increase the likelihood of false stress locations.

Examination of the detailed stress patterns in the various database subsets shows a few specific causes for many missed stresses and false locations. For example, only subsets 1C, 1D, 4C, and 7B had less than 90% of all perceived stresses correctly found. In subsets 1C and 1D, 10 of the 33 stressed syllables that were not located were in utterance initial position. An error in the STRESS program occasionally caused such initial stresses to be missed, and stress to be placed on the second syllable even though it had less Fo rise and nucleus duration. Twelve (30%) of the 40 missed stresses in subsets 1C, 1D, and 4C were for the word "young", while another ten (25%) were for the second syllable on "enroll".

Sixty nine (29%) of all the 238 false stresses located in the full set of 255 sentences were cases of the auxiliary verb "will" (or, where the two syllables are erroneously detected as one, then the sequence "will en-") being called stressed. All but one of the eight cases of the auxiliary verb "are" in subset 1E were falsely located as stressed. Subset 3D stands out as having many false stress locations. This is the one subset where voiced and unvoiced obstruents are very frequent, and it is the presence of such obstruents that appears to cause all or almost all of the false stresses in that subset. Other prominent sources of false stresses included: repeated words in coordinate constructions (36, or about 56% of the false alarms in subsets 8A, 8H, and 8K); conjunctions (12, or 19% of the false alarms in subsets 8A, 8H and 8K), and relative pronouns (8, or 35% of the false alarms in subset 11A). While many of these words were not perceived as stressed, they showed many of the prosodic correlates of stressed syllables. Perhaps these are instances of the listeners hearing a stress level which is dictated more by expectations determined by the sentence structure than by acoustic information. One other possible cause of

27 false stresses involves utterance-final or prepausal syllables. There is a special test in the STRESS program that locates prepausal stresses if the prepausal syllables are of sufficient duration. Since the durations of nuclei are now usually defined longer than before (see pages 16 to 18), unstressed prepausal syllables are more prone to be erroneously detected as stresses.

In summary, the stress location results were very good, though there is room for improvement. This is evident when one considers that 18% of the 1809 syllables detected by the syllabification routine were either perceived as stressed and not located as stressed, or perceived as non-stressed and located as stressed; that is, 18% of the syllables were <u>confused</u> between perceived and automatically detected stress levels. This is considerably more than the 3 to 6% confusions in perceived stress levels from trial to trial or from listener to listener (Lea, 1976c, pp. 27-29). An ideal stress location algorithm might be expected to exhibit around 5% confusions when compared with perceived stresses. Part of that remaining 13% or so might be eliminated by better syllabification results and by correcting the current errors that miss utterance-initial stresses and introduce erroneous utterance-final (prepausal) stresses. Other improvements could come from adjustments of syllabic durations, Fo values, and energies on the basis of the vowel identity (or formant Fo values; see Lea, 1976c, p. 19) and phonetic context. Perhaps totally different ways of combining Fo, energy, and duration cues could improve stress location scores, though other published algorithms have not performed better (Sargent, 1975; Cheung, 1975).

3.5 Testing Intonational Hypotheses

The success of the archetype algorithm for stressed syllable location shows the following major features of the intonation of English sentences:

- Fo rises substantially at the first stress in each major phrase;
- Fo falls gradually after peaking near or somewhat after the first stress in the phrase;
- Fo rises slightly at all other stresses in the phrases (but usually not so much as to yield a fall-rise Fo valley within the phrase).

We also noted the following complicating effects of phonetic structure on on Fo contours:

- Fo dips during voiced obstruents, and
- Fo is high immediately after unvoiced obstruents, and then rapidly drops to values dictated by the stress and other large-unit intonational effects.

Fundamental frequency contours in the 255 sentences also show several other types of intonational regularities besides Fo valleys at phrase boundaries and increases of Fo at stresses. For example, we noted in the previous section that 99% of the initial stresses in sentences were coincident with (or immediately before) the Fo peak in the sentence. Turning this around, we may note more generally that

- Fo rises steadily in the initial part of sentence, until the first stress, where it peaks.

This regularity was noted at least as early as 1929 (Armstrong and Ward), but is sometimes obscured in arbitrary sentences by effects such as phonetic influences on Fo causing fall-rise valleys before the first stress, or early brief jumps of Fo, after unvoiced obstruents, to values just higher than the values in the initial stress. Sometimes a syllable after the first stress can have a brief Fo peak above that in the first stress, due to an unvoiced obstruent. However, in the 255 sentences analyzed, such effects did not mask this regularity. The all-sonorant sentences obviously had no such complications, and I found that the phonetic effects in the sentences with obstruents could be eliminated by disallowing the first two Fo points after a period of unvoicing from defining the peak, and simply stating that the first stress is the syllable whose syllabic <u>peak</u> immediately precedes the Fo peak. In a couple of borderline cases of stress, where some of the listeners heard the utterance initial syllable as stressed, and the computer program located the syllable as stressed, the Fo peak appeared there even though the majority of listeners didn't perceive the syllable as stressed. These seem so much like initial stresses that they shouldn't be taken as refuting the regularity of coincidence of Fo peak and initial stress. Also, there were two cases where the Fo peak did not align with the first stress because an emphasized syllable later in a sentence caused an unusually high Fo in its region, which exceeded the Fo peak on the initial stress. However, in both cases, this later Fo peak might be ruled out either by setting a maximum time (after voicing onset) before the first stress must be encountered

or else by noting the presence of a large Fo valley before that delayed Fo peak. This effect may be stated specifically as an intonational hypothesis:

- An emphasized or contrastively stressed syllable in a sentence will have an unusually high peak Fo value that can equal or exceed that of the initial stress in the sentence.

Much more testing would be needed to verify this effect of emphasis and contrastive stress.

Another intonational regularity that was verified firmly with the 255 sentences was the following:

- Fo falls after its highest Fo value in the last stress, to a low value at the end of each declarative, command, and WH question. Fo dips, then rises, within the last stress of yes/no questions, and rises throughout subsequent unstresses.

This was found to be true for every sentence except two: one declarative which the talker spoke with a sense of being incomplete ("Men will know...." [KNOW WHAT?]) and a yes/no question with emphasis on a quantifier, so that it seems more like a WH question of "How many" than a yes/no question ("Will all your men know?") With over 99.5% of the declaratives, commands, and WH questions, and about 95% of the yes/no questions, satisfying this terminal contour regularity, we can consider it well verified.

So, we now have Fo rising to the first stress, and (for all but yes/no questions) falling from the last. What happens between the first and last stress? A preliminary study of the first 58 of the sentences showed that, in 91% of the cases:

- Fo falls from one stress to the next in a sentence (or clause).

The exceptions were spanning major syntactic boundaries that were followed by highly stressed syllables. Another regularity was that:

- Fo on unstressed syllables is lower than on all preceding stresses, and is usually at or below a value along the line between the values of the immediately preceding and following stresses.

Exceptions were when an unstressed syllable had higher Fo value than the preceding (phrase-initial) stress because the stressed syllable was short and Fo continued to rise, plus the unstress was a high vowel while the preceding stressed vowel was low and hence had an intrinsically lower Fo (examples: "any," "many").

I expected from published claims that Fo might mark subordination of one phrase under another, but found no clear regularity. I also was unable to simply characterize any unique Fo contours in coordinate NP's or other coordinate constructions (cf. Lea, 1972).

Table VIb summarizes some prosodic cues that were found for one clear structural contrast; namely paranthetical or appositive (non-restrictive) relative clauses. The paranthetical (whose description and prosodic values are to the left of the slashes in Table VIb) is preceded and followed by longer time intervals between it and the surrounding syllables; that is, by a form cf brief pauses. Fo falls dramatically before the paranthetical, and, after the paranthetical, rises substantially. A Tune II rise in Fo marks the incompletion and interruption the paranthetical produces. There do seem to be two distinctive types of parantheticals, though; one for which the Tune II occurs <u>before</u> the paranthetical (L111 and L112) and one for which it occurs at the <u>end</u> of the paranthetical (L117, L120, L121). Thus it appears that:

- Paranthéticals are demarcated by large Fo variations and long intersyllabic time intervals at their boundaries, and by Tune II Fo contours.

Finally, a very clear example of the potential of using Fo contours to detect syntactic structures is shown for the sentences of the "Flying Planes Paradigm" (named after the two alternative structures of the ambiguous sentence "They are flying planes."). Sentences like "Lawmen are lying men", with the structure NP-Copulative-ADJ+N, have Fo valleys (hence, phrase boundaries) only before the ADJ. In contrast, sentences like "Lawmen are ruling Maine.", of the structure NP-AUX-V-N have valleys (boundaries) before the V and the N. This was always true for all the sentences in subset 1E. Thus, Fo contours, and Fo-detected boundaries, can clearly distinguish between alternative phrase structures. It is interesting to note that no such contrast was evident in perceived stress patterns (Lea, 1976c p. 42).

Table VIb. Prosodic Cues to the Presence of Parantheticals

	BOUNDARIES NON-PARANTHETICAL/PARANTHETICAL	INTERSYLLABIC TIME INTERVAL	Fo FALL (FROM PREVIOUS STRESS)	TUNE	Fo RISE (TO NEXT STRESS)
L107/L111	Men/Lynn,-who knew	28/48	21/30	I/II	14/29
	Ron-ran Maine.	39/40	18/33	I/I	7/18
L108/L112	Men/Lynn,-whom Ron	30/48	21/27	I/II	10/18
	knew/knew,-ran Maine	26/41	19/36	I/I	6/18
L110/L117	Men/Men,-Ron	30/68	17/20*	I/I	8/3*
	knew/knew,-ran Maine	30/52	20/7*	I/II	7/16*
L118/L120 /L121	men/...May,-whom /...May,-whom	28/80 /40	12/26 /20	I/II /II	8/20 /18
L123/L124	...oil/...oil,-- -Wayne/which Wayne...	41/74	12/11	I/I	11/14

* No stress is in the paranthetical of L117. The Fo fall as given is from "Mon" (+5) to valley, the rise is from the valley to "Ron" (+1), the next fall is from the peak in "knew" (+2) to the second valley which is also within "knew", next rise is from the second valley to "ran" (+3), neglecting a local Fo jump due to a glottal stop.

We may summarize by saying that many intonational regulations could be noted with the controlled contrasts in the 255 sentences, and they provide information that might be useful in speech understanding systems.

3.6 Acoustic Correlates of Stress

The stressed syllable location program does a fairly good job of utilizing some major acoustic correlates of stress. However, our acoustic prosodic analysis of the 255 sentences provides extensive data for further studies of the acoustic correlates of stress. The Fo contour can provide peak and average Fo values in each syllable, Fo contour slopes and shapes within each syllable nucleus, and more global Fo contour features. The sonorant energy function and syllabification procedure provide the duration of the syllabic nucleus, the peak energy value in the nucleus, a measure of the energy integral for the nucleus, and other energy and duration information. Further study of such features can and should be done. In particular, such studies may help devise better algorithms for stressed syllable location. Unfortunately, time did not permit our studying such data in any detail.

Further studies can also include investigations of how stress decisions can be adjusted to take account of the intrinsic prosodic features (phonetically-dictated energies, durations, and Fo values) of various vowels and consonants. We know that an unstressed high vowel may have higher Fo than a stressed low vowel, while an unstressed low vowel may have higher energy and longer duration than a stressed high vowel. Voiced consonantal contexts also cause a vowel to be longer, its Fo to be somewhat lower, and its energy to be somewhat higher. Further studies could perhaps make more or better use of <u>relative</u> values of prosodic features in comparing one syllable with its neighbors.

The controlled contrasts in the 3300-sentence speech data base should be very useful for undertaking such future studies of acoustic correlates of stress.

3.7 Specific Implications for Speech Understanding Systems

In summary, the studies of acoustic prosodic patterns in the 255 sentences have provided important confirmation of our procedures for phrase boundary detection, syllabification, and stressed syllable location. They also have

suggested specific ways in which such algorithms might be improved. We thus have the promise of improved prosodic analysis tools that may be useful in speech understanding systems.

More importantly, these studies have firmly established various prosodic regularities that may be used to predict prosodic patterns accompanying hypothesized sentence structures within speech understanding systems. Regularities of syllabification, automatic stress assignment, boundary placement, intersyllabic time intervals, and Fo contour shapes may be used to adjust the scores on hypothesized words or word sequences, in a manner similar to that we developed for the BBN HWIM system. For example, we may predict that words like "worry" or "moral" may be detected as monosyllabic, and not penalize a word match that involved only one acoustically-detected syllable for such theoretically-multisyllabic words. We may allow "will en-" in a sentence like "Ron will enroll airmen." to be found as one syllable, or we may first try to improve syllabification to find the missing syllable boundary (by using a new spectrally-weighted energy function). From our studies we may predict more precisely just which syntactic structures will exhibit Fo boundaries, and where they will be positioned. Then, if we find agreement with such predictions, we can reward that structure with a higher priority in the hypothesizing and testing. Similarly, if stresses occur on the wrong syllables for a certain structure (that is, they are not those predicted from previously observed regularities), we could decrease the priority of hypothesizing that structure.

Specific structural features, such as the sentence being a yes/no question, a paranthetical being present, or one of alternative syntactic bracketings being possible, can also be gleened from the prosodic data, to aid parsing and the overall control strategy of a speech understanding system. Much has been learned from the study of the controlled speech texts, and much more could be learned from extensive further studies.

4. REVIEW OF PROSODICS RESEARCH PROGRAM

Sperry Univac's research on prosodic guidelines to speech understanding produced many experimental results that help us better understand how prosodic structures relate to other aspects of linguistic structures, such as phonemic sequences and phrase structures. We have presented theoretical arguments about the need for extracting from the acoustic speech signal some prosodic cues to the large-unit linguistic structure, without dependence upon the prior determination of phonemic structure and recognition of the words in the sentence (Lea, Medress, and Skinner, 1972a). Vital assumptions of a prosodically-guided approach to speech understanding have been verified from a variety of experiments. We thus have both theoretical and experimental reasons for promoting the use of prosodic information in speech understanding systems. These will be outlined in section 4.2, following a tabulation in section 4.1 of all major results from our research for ARPA. Section 4.3 provides a review of our various cooperative efforts to make prosodics an important aspect of speech understanding systems.

In sections 4.4 to 4.7, we briefly review four major areas of prosodic studies: intonation and phrase boundary detection (4.4), perceived stress patterns (4.5), automatic location of stressed syllables (4.6), and timing cues to linguistic structure (4.7). We also have made some initial attempts to define and test some specific procedures for using prosodic information in speech understanding systems (4.8). A major contribution suitable for aiding future studies is our development of a large speech database with controlled linguistic contrasts from sentence to sentence (4.9).

4.1 Overview

Table VII on pages 41 to 43 summarizes the major contributions that have come from Sperry Univac's ARPA-sponsored research. We have been leading advocates of the use of prosodics in speech understanding, doing what we can to precisely define the role of prosodics (as listed in section A of Table VII). We have also cooperated with other ARPA/SUR contractors in many general aspects of the overall large speech understanding programs (section B of Table VII), and have developed and provided computer programs and other services to system builders (section C of Table VII). Section D of Table VII (pages 42 and 43) shows that we contributed in essentially all aspects of prosodics: intonation, stress patterns, phonetic durations, rhythm, pauses, rate of speech, acoustic correlates of stress and

TABLE VII. CONTRIBUTIONS OF SPERRY UNIVAC TO THE ARPA SPEECH UNDERSTANDING RESEARCH PROGRAM (1972–1976)

A. DEFINING THE ROLE OF PROSODICS IN SPEECH UNDERSTANDING

- Stressed syllables are important because they occur in important words, exhibit close phonemic-phonetic correspondence, are more carefully articulated and more reliably analyzed phonetically, are good indicators of syntactic structures, and are closely associated with predictable phonological distortions at various rates of speech.

- Various machine transcriptions of speech (i.e., results of automatic segmentation and labelling of speech) were analyzed, and it was shown that far fewer errors in vowel and obstruent classification occurred in stressed syllables than in unstressed or reduced syllables.

- Linguistic and perceptual arguments suggest that syntactic structures, detectable from prosodic patterns, should be used at early stages of speech understanding.

- Sentences which had been troublesome to the BBN speech understanding system were processed through the Sperry Univac prosodic analysis programs, and specific prosodic cues were found that could be used to determine the type of sentence and the specific syntactic bracketing intended by the talker.

- An overall strategy for prosodically-guided speech understanding has been specified. It involves use of stressed syllables as anchor points for reliable phonetic and phonemic analysis, restricting expensive acoustic analyses to those areas where prosodics say such analysis is needed, guiding the selection of applicable phonological rules, and detection of aspects of syntactic structure directly from prosodic patterns.

B. COOPERATIVE EFFORTS WITH OTHER ARPA/SUR CONTRACTORS

- Our syntactic and prosodic analysis of 250 sentences produced by SUS contractors resulted in the selection of the "31 ARPA Sentences", used in various common studies such as workshops on parameterization, speech segmentation, and phonological rules.

- Sperry Univac and other ARPA contractors have co-operated on major common tasks of selecting speech data bases, standardizing recording procedures and phonemic notations, compiling and applying sound structure rules, comparing speech parameterization techniques and speech segmentation results, and other comparative activities. In particular, a tutorial was presented on prosodic structures, prosodic information on selected data bases was supplied to several work-shops, sample parameters and segmentation results were presented at the respective workshops, and sessions on prosodic structures were chaired at workshops on phonological rules.

- Sperry Univac actively participated in steering committee meetings and other activities guiding the overall SUR program. Dr. Mark Medress of Sperry Univac served as Assistant to the Chairman, and later as Acting Chairman of the Steering Committee.

- Sperry Univac was actively involved in the development of ideas for a five-year follow-on program to extend the current five-year SUR program.

- Sperry Univac produced nine semi-annual reports, five other ARPA-sponsored reports, 5 journal papers, 14 oral presentations and 25 SUR NOTES describing our research (see Appendix B), and extensive communications over the ARPANET.

C. COMPUTER PROGRAMS, AND APPLICATIONS TO ARPA/SUR SYSTEMS

- Computer programs were developed and supplied to ARPA/SUR contractors, providing the following prosodic information:

  - Fo Contours
  - Intonational Phrase Boundaries
  - Syllabification
  - Stressed Syllable Locations

  These programs were implemented in the BBN system, and used for devising similar programs at SDC.

- A procedure has been developed for using prosodically-detected phrase boundaries to weigh word and phrase hypotheses in the Bolt Beranek and Newman (BBN) HWIM speech understanding system. The state-transition arcs of the augmented transition network grammar were specially marked if they were expected to be immediately preceded by intonationally-detected phrase boundaries. The scores on words associated with the arcs were increased if expected boundaries were detected, or decreased if expected boundaries were missing in the acoustic-prosodic data. Sixteen BBN sentences were processed through a computer program that detected phrase boundaries at fall-rise valleys in fundamental frequency contours. Analysis of simple traces of the hypothesizing, testing, and constructing of syntactic stru ures by the HWIM system showed that prosodic adjustment of scores would increase the likelihood of correct words and phrases being selected before incorrect ones. These ideas were later refined, tested further, and implemented in the HWIM system, but time didn't permit their full testing.

TABLE VII. CONTRIBUTIONS OF SPERRY UNIVAC TO THE ARPA

SPEECH UNDERSTANDING RESEARCH PROGRAM (1972–1976)

D. EXPERIMENTAL RESULTS

Perception of Prosodies

- Listeners can reliably perceive which syllables are stressed (with 5% confusion).
- Perceived stress patterns agree with those assigned when subjects are given only the written text.
- Certain words ("content words") are consistently perceived as stressed, while others ("function words") are perceived as unstressed or reduced.
- Repeated verbs or nouns in coordinate constructions have lower perceived stress levels than in simple constructions.
- Verbs, auxiliary verbs, and conjunctions have lower perceived stress levels when in subordinate phrase structures.
- Listeners can reliably perceive phrase boundaries in spectrally-inverted speech.

Intonational Phrase Boundaries

- Substantial Fo valleys occur at major phrase boundaries (before NP's, V's, ADV's, PP's, Clauses).
- About 80-90% of the major phrase boundaries in speech can be detected from Fo valleys.
- False boundary detections result from Fo variations near obstruents.
- The Fo-detected phrase boundary occurs just before the first stress in the following phrase.

Fo Contours

- Fo contours are a superposition of fall-rise clause contours, archetype phrase contours, Fo rises at stress positions, and Fo variations at obstruents (i.e., dips during voiced obstruents, and sudden jumps at unvoicing with subsequent rapid fall from high values).
- Fo contours within phrases, and phrase boundary breaks in the Fo contour, may be modelled by a statistical curve fitting procedure using a modified form of the gamma distribution.
- The Fo Peak in a sentence is at the first stressed syllable of the sentence.
- Fo falls after the last stress of all declaratives, commands, and WH questions, and rises within and after the last stress of yes/no questions.
- Succeeding stresses have progressively lower Fo values, except at major syntactic boundaries, where Fo may rise for highly stressed or emphasized syllables.
- An unstressed syllable has lower Fo than all stresses that precede it within a clause.
- Parenthetical phrases are preceded and followed by long intersyllabic intervals, and marked by a Tune II rise, and large Fo variations at both ends of the parenthetical.
- Subordination of phrases does not appear to be readily detected from Fo contours.
- Glottal stops, detectable by large local variations in Fo contours, are twelve times more likely to occur before a stressed than an unstressed vowel in all-sonorant phonemic sequences. They also frequently mark phrases boundaries.

Syllabification

- Over 90% of the syllables in connected speech may be found from high energy nuclei surrounded by dips of 4db or more in energy.
- The beginning and ending of a syllabic nucleus may be quite accurately located at the outermost points where energy is at least half way above the dip in energy, towards the peak energy level in the nucleus.
- Syllables are not detected when, during all-sonorant sequences, the energy level does not dip adequately for syllable boundary detection. Another spectrally-weighted energy function or segmental information (such as formant transitions and automatic detections of non-vowel sonorants) might help locate such missing syllable boundaries.

Stressed Syllable Location

- Over 90% of the stressed syllables in connected speech may be located at those syllabic nuclei that have non-falling Fo and are highest in energy in the vicinity of either (a) the Fo rise to the peak value at the beginning of a phrase, or (b) local Fo rises about a gradually falling archetype line in the later part of a phrase.
- About 20% of all stress locations are false (not pointing to perceived stresses), due to selection of the wrong nucleus in a neighborhood of an Fo rise. Adjustments of durations, intensities, and Fo contours for vowel height and consonantal context may reduce such errors.
- Stressed syllable locations from using Fo rises alone or long-duration nuclei alone were found to be considerably less accurate and produce more false stresses than locations using the above "archetype contour" algorithm.
- The first stress in a sentence was found 99% of the time by locating the nucleus immediately preceding the peak Fo in the sentence.

TABLE VII. CONTRIBUTIONS OF SPERRY UNIVAC TO THE ARPA SPEECH UNDERSTANDING RESEARCH PROGRAM (1972–1976)

D. EXPERIMENTAL RESULTS (continued)

Isochrony of Stresses

- Time intervals between stresses are a linear function of the number of intervening stresses.
- Interstress intervals tend to cluster near about 0.4 seconds (i.e., tend toward isochrony of stresses) primarily because of the alternating stress/unstress pattern of English.

Structural Pauses

- Long periods of unvoicing (including periods of silence) occur between clauses.
- The duration of a pause is usually one rhythm unit (one average interstress interval) between clauses and two units between sentences.
- Hesitation pauses are usually substantially longer than structural pauses.

Phrase-Final Lengthening of Vowels and Sonorants

- Vowels and sonorant consonants are substantially lengthened in phrase-final positions.
- The phrase-final lengthening extends to groups of neighboring syllables including the last stress in a phrase and any subsequent unstresses up to the first stress in the next phrase plus, in some cases, the next earlier stress and neighboring unstresses. Hence detected phrase boundaries (at the end of the lengthened group) did not always occur at the time of the syntactic boundary.
- Over 90% of the phrase boundaries perceived by listeners who heard spectrally inverted speech were detectable from groups of lengthened syllables

Interstress Intervals as Phrase Boundary Cues

- Over 95% of the major phrase boundaries perceived in spectrally inverted speech may be detected from long interstress intervals ($\geq$ 0.5 seconds) spanning the boundaries.

Rate of Speech and Phonetic Distortions

- The time interval between two stresses was shown to be inversely correlated with the percentage of phones (between those stresses) that had been erroneously categorized by automatic labelling schemes. The interstress interval was demonstrated to be a better predictor of error rate (and, thus, a better indicator of applicable phonological rules) than other measures of speech rate, such as the number of syllables per second.

Prosodic Hypotheses

- A careful study of the literature and previous analyses of prosodic data resulted in a compilation of an extensive set of hypotheses and rules relating prosodic patterns (intonation, stress, rhythm, etc.) to linguistic structures (sentence types, syntactic bracketing and syntactic categories, phonetic sequences, semantic structures, etc.).

Database of Controlled Linguistic Contrasts

- A data base of 1100 sentences was designed to carefully isolate factors influencing prosodic and phonetic structures. A set of 178 "Phonetic Sentences" is especially suitable for testing automatic schemes for formant tracking, phonetic segment classification, and phonological rules application. A set of 922 "Prososyntactic Sentences" was designed such that various minimal pairs of sentences could isolate prosodic effects due to sentence type, syntactic bracketing, subordination, coordination, lexical stress patterns, semantic contrasts, and phonetic sequences. The data base includes sentences typical of those handled by the ARPA speech understanding systems.
- The database has been divided into small subsets that test specific prosodic hypotheses and linguistic contrasts. Initial subsets totalling 255 sentences have been digitized and processed through prosodic programs to study such regularities as the placement of prosodically detected phrase boundaries as stresses move, which phrase boundaries are marked in Fo contours, perceived and automatically detected stress patterns, and overall Fo contours. Many further tests could be undertaken with this database.

boundaries, listeners' perceptions of prosodics, syllabification, and the compilation of data bases and hypotheses for extensive controlled investigations of prosodic structures. Contrary to much of past work with prosodics and some current work in speech synthesis, which needs to focus only on the salient features of acoustic prosodic data (cf. e.g., Allen and O'Shaughnessy, 1975; O'Shaunessy, 1976), these studies have been very exacting, with use of actual acoustically-derived information such as computed fundamental frequency contours (with all their inexactitudes, occasional octave errors, local perturbations, etc.), sonorant energy functions, and automatic voicing decisions. Our computer programs for prosodic analysis then directly use that imperfect information, such as syllabification being based on sonorant energy contours, and stressed syllable location being based on Fo and energy contours as well as the syllabification results.

Basically, our effort has been two-pronged: (1) experimental research about prosodic structures, and (2) cooperation with other ARPA contractors for general tasks and for development of prosodic aids to the developing speech understanding systems. In sections 4.4 to 4.9 we will review the experimental research. In sections 4.2, and 4.3, we will review the cooperative efforts in speech understanding system development.

4.2 Defining the Role of Prosodics in Speech Understanding Systems

Prior to the ARPA/SUR program, prosodic cues to sentence structure, and prosodic aids to the location of reliable acoustic phonetic information, were given little or no attention in speech recognition efforts. The strong motivations for the use of prosodic patterns in speech recognition procedures were thus presented in some detail in our first report (Lea, Medress, and Skinner, 1972a, section 2), and subsequent reports (notably, Lea, 1976b). In particular, we showed that stressed syllables are of prime importance in speech recognition, because of: (a) the occurrence of stressed syllables in semantically important words; (b) the close correspondence between detected phonetic structure and underlying phonemic structures in stressed syllables; (c) the much higher reliability of phonetic classification possible in stressed syllables (as evidenced by the analysis of results from the CMU Speech Segmentation Workshop); (d) the vital cues to syntactic structure that stressed syllables provide; and (e) the close association between time intervals between stresses (as rate-of-speech measures)

and applicable phonological rules.

In our first progress report (Lea, Medress, and Skinner, 1972a), we reviewed linguistic and perceptual arguments that prosodic structures should be used to detect aspects of syntactic structure independently of any phonemic analyses and word matching algorithms. Linguistic arguments suggest that phonetic sequences are not invariant linear strings that occur each time a word is spoken, and, taken alone, they cannot be relied upon to provide all the information needed for determining the word sequence in a sentence. These arguments have been vividly verified by the demonstrations with the 1976 ARPA speech understanding systems (especially at CMU), which showed that syntactic constraints were very important in providing successful speech understanding. Perceptual arguments indicate that human listeners, as successful archetypes of speech understanding mechanisms, use phrase units at earliest stages of speech perception, and that those phrases are detected from prosodic information. Our experiments have clearly confirmed the marking of large-unit linguistic structures in prosodic patterns.

To further confirm the value of prosodics in speech understanding systems, we conducted two studies, which, though they are experiments and thus might be listed under section D of Table VII, have as their primary consequence the demonstration of the valuable role of prosodics in speech understanding, and thus are listed under section A of Table VII. We investigated various machine transcriptions of speech resulting from the automatic segmentation and labelling of speech provided by several research groups reporting at the 1973 CMU Symposium of Speech Segmentation. We found that far fewer errors in vowel and obstruent classification occurred in stressed syllables than in unstressed or reduced syllables. In another study, sentences which had been troublesome to the BBN speech understanding system were processed through our prosodic analysis programs, and specific prosodic cues (F_0 contours, detected phrase boundaries, stressed syllable locations, pauses, and timing cues) were found that could be used to determine the type of sentence and the specific syntactic bracketing intended by the talker.

We also defined an overall strategy for speech understanding which uses stressed nuclei as islands of phonetic reliability to be detected in early stages of phonetic analysis, and uses prosodically-derived syntactic hypotheses to guide syntactic parsing and an overall analysis-by-synthesis process.

It is interesting that, after our efforts to clearly define an important role for prosodics in speech understanding systems, the Steering Committee of the ARPA Speech Understanding Research program, in its mid-term review of the total SUR program, suggested that a major attack should be mounted on the area of prosodics, since this source of knowledge had not been used in any previous system, though it offered the possibility of a unique contribution to sentence disambiguation and overall system control strategies.

4.3 Cooperative Efforts to Advance the Development of Speech Understanding Systems

Besides providing solid arguments for the use of prosodics in speech understanding, we have participated in various other aspects of system development. Under the guidance of the Steering Committee, Sperry Univac has been engaged in a variety of cooperative efforts to aid the progress of the overall SUR program. Our syntactic and prosodic analysis of 250 sentences produced by the system building contractors resulted in the selection of 27 generally interesting sentences representive of the task domains being used in the various systems. These ultimately formed the bulk of the "31 ARPA sentences" (Lea, Medress & Skinner, 1973b) used in various common studies such as workshops on parameterization, segmentation, and phonological rules.

Sperry Univac and other ARPA contractors have cooperated on major common tasks that were helpful to the various systems and the research being conducted within the program. These included: selecting speech data bases; standardizing recording procedures; comparing speech parameterization techniques; developing the uniform ARPABET phonemic notation (Medress, 1972, SUR Note 32); comparing various methods of speech segmentation and labelling; and compiling, comparatively evaluating, and applying sound structure (phonological) rules. At an early stage in the program, Lea presented a tutorial on prosodic features and linguistic structures, at the ARPA Seminar on Acoustic Phonetic Characteristics of English Sentences. He also chaired sessions on prosodic phenomena at the workshops on phonological rules.

Mark Medress of Sperry Univac served during part of the SUR program as Assistant to the Chairman, and later as Acting Chairman, of the ARPA/SUR Steering Committee. Sperry Univac also was very active in other Steering Committee activities, including proposing specific ideas and plans for a five year follow-on program to extend and apply the results of the five-year SUR program.

A concrete product of Sperry Univac's work was the circulation of nine regular ("semiannual") progress reports, five other ARPA-sponsored reports, five journal papers, 14 oral presentations, and 26 SUR Notes.

While these various cooperative efforts within the SUR program required a significant portion of our effort and other groups' efforts, they represent one of the greatest benefits of the ARPA/SUR program. Little had been done before this program to compare alternative methods in speech analysis or to encourage close cooperation and even direct competition among speech research groups. Interchange about parameterization techniques, speech segmentation procedures, phonological rules, syntactic models, semantic and pragmatic constraints, and system structures has significantly contributed to the success of the program and the specific systems. The compilation and application of phonological rules was a major contribution to speech sciences. It is worth noting, by the way, that many of the selected phonological rules depend upon prosodic information such as stress patterns.

Perhaps the most specific contributions of Sperry Univac's work to the development of speech understanding were in providing prosodic analysis routines and specific proposals of how to use prosodics in the ARPA/SUR systems. We developed and circulated to the ARPA/SUR community, a FORTRAN program for obtaining an Fo value every 10 ms, based on a center-clipped autocorrelation analysis (Skinner, 1973a,b). This program was subsequently modified and used by other ARPA/SUR contractos, including BBN and SDC. Another program which was delivered to SUR contractors was the FORTRAN program "BOUND3", which detects syntactic boundaries from fall-rise valleys in Fo contours and long periods of unvoicing ("pauses"). This program was incorporated into the BBN HWIM system, and ideas from it were also used at SDC.

A third FORTRAN computer program (CHUNK") used sonorant energy contours to locate the peaks, beginnings, and endings of syllabic nuclei, and to locate syllable boundaries. This program was incorporated into the BBN HWIM system, and a similar program, based in part on Paul Mermelstein's work at Haskins Laboratories, was incorporated into the SDC system. Our FORTRAN program for locating stressed syllables ("STRESS") was based on archetype Fo contours in detected phrases and high values of energy integral in syllabic nuclei. This program was incorporated into the BBN HWIM System.

Last, but by no means least, of our efforts to apply prosodics to SUR systems was our development of a procedure for using Fo-detected phrase boundaries to adjust the scores on word and phrase hypotheses in the Bolt Beranek and Newman (BBN) HWIM speech understanding system, so that correct words and structural hypotheses will be proposed at earlier stages in parsing, and erroneous theories can be avoided. The state-transition arcs of the augmented transition network grammar were specially marked if they were expected to be immediately preceded by intonationally-detected phrase boundareis. The scores on words associated with the arcs were increased if expected boundaries were detected, or decreased if expected boundaries were missing in the acoustic-prosodic data. Sixteen BBN sentences were processed through the computer program that detected phrase boundaries at fall-rise valleys in fundamental frequency contours. Analysis of sample traces of the hypothesizing, testing, and constructing of syntactic structures by the HWIM system showed that prosodic adjustment of scores would increase the likelihood of correct words and phrases being selected before incorrect ones. These ideas were later refined and modified to handle BBN's new shortfall density scoring procedure, tested further, and implemented in the HWIM system. BBN researchers planned to test the HWIM system with and without prosodic information but were unable to perform such tests before their contract ended. Still, as was noted in section 2.2, our studies showed that prosodics caused a rearranging of the priorities of hypotheses such that correct theories would have been tried earlier than without the prosodic guidelines, so that false parsing paths could be avoided, parsing could be more efficient, and more correct parses should result.

4.4 Intonation and Phrase Boundaries

An algorithm was devised for segmenting speech into grammatical phrases, by marking phrase boundaries at the bottoms of "substantial"[1] fall-rise valleys in fundamental frequency (F_0) contours. This algorithm was implemented as a FORTRAN program on the Sperry Univac interactive speech research facility, then supplied over the ARPANET to all SUR contractors. It uses Fo data obtained from the Sperry Univac fundamental frequency tracking program (Lea, Medress, and Skinner, 1973a, Appendix A). The algorithm also successfully detected clause and sentence boundaries wherever long (350 millisecond) stretches of unvoicing (i.e., "pauses") occurred.

1. In earlier studies, a "substantial" Fo valley was considered to be defined by a minimum of 7% fall and 7% rise in Fo. In the most recent studies, we used 4 eighth tones as the threshold value for an Fo valley.

A series of "natural experiments" (cf. Anderson, 1966) were conducted to test the algorithm. In such "natural experiments", one does not directly control an independent variable (such as syntactic bracketing) and study resultant changes in a dependent variable (such as valleys in Fc contours); rather, he simply looks at the data obtained from naturally-occurring phenomena (such as the speech which had previously been recorded at Purdue University and identified as the Rainbow Script, spoken by six talkers, and the Monosyllabic Script, spoken by two talkers). Our first experiment included those texts, plus 13 of the 31 ARPA sentences. For such speech texts, we demonstrated that over 80% of all intuitively predicted syntactic boundaries were detected from substantial fall-rise valleys in Fo contours. Over half of the "missing" boundaries were between noun phrases and auxiliary or main verbs. In some later reports we excluded such NP-AUX and NP-V boundaries, getting scores nearer 90% for all other boundaries.

Some "extra" boundaries were detected at places in the syntax where they had not been expected, and some "false" boundaries also were detected where they obviously had no relation to syntactic structures. The false boundaries were almost all due to local Fo variations introduced by obstruents. The "extra" boundaries and "missing" boundaries (expected but not detected) needed to be better understood, yet the uncontrolled nature of the speech texts made it difficult to find simple explanations. More controlled studies with sentence pairs with minimal differences in structure needed to be conducted.

We extended those tests in a later study, to include the full set of 31 ARPA man-computer interaction sentences. Boundary detections were somewhat more reliably found with speech read from a written text than in some of the simulated man-computer interactions. This test still involved uncontrolled speech texts.

In 1975, after the large Sperry Univac speech database had been designed and recorded, tests were conducted on the ability to detect phrase boundaries in a subset of 159 designed sentences, involving three talkers. All the sentences in this subset were simple (unembedded) declarative sentences with one of six phrase structures. The majority were of the form "Ron will enroll NP", to test how the Fo-detected boundary before the NP moves as the first stress in the NP moves. Unfortunately, two of the talkers showed very little Fo variation throughout each utterance, so that we could not conclusively determine <u>which</u>

syntactic constituents are separated by fall-rise patterns of fundamental frequency. The main conclusion that resulted from this study was that the rise in Fo after any detected boundary will begin at the first stress in the following constituent (for all the talkers). The Fo-detected boundary invariably occurs just before the first stress in the following phrase. Of course, other factors like dips of Fo during nearby voiced obstruents could cause local movement of the boundary, but the syntactic structure and stress patterns dictate that the boundary be just before the first stress. In all-sonorant sentences such as the 159 studied then, the placement of the boundary immediately before the first stress of the next phrase becomes very apparent.

The most recent study of intonational phrase boundaries, described in section 3.3 of this report, further verified this placement of the Fo boundary. However, with the one talker used in this latest study, whose Fo contours are more animated and show clear boundary markings, we were able to also determine <u>which</u> constituents are regularly accompanied by Fo-detected boundaries. After a stressed constituent, we find that noun phrases, sentence adverbs, conjuncts, relative clauses, and paranthetical are preceded by Fo boundaries, as are stressed main verbs (and auxiliary verb phrases if and only if they contain a stress such as a negative). The two parts of a compound noun are also separated by an Fo boundary. Since some words, like main verbs, lose their stress in some constructions (e.g., in coordinate constructions and subordinate phrases), those words will not be preceded by Fo boundaries in such positions.

Boundaries were found in this latest study to occur more regularly when unstressed syllables intervene between the last stress of the previous phrase and the first stress of the following phrase. In general, in predicting where Fo boundaries should occur, expected stress patterns should be taken into account, so that boundaries will not be expected in coordinate structures with repeated words, or in some subordinate structures. Even the presence or absence of unstresses between stressed syllables could be used to refine the probability of detecting a phrase boundary. This study indicated that boundaries should perhaps be expected at the first stress after the auxiliary verb in a yes/no question, between a noun phrase and the relative pronoun of its subordinate relative clause, and perhaps even between an adverb and the adjective it modifies within a noun phrase.

The minimal contrasts in structure within pairs of the database sentences have been useful in highlighting these refinements in boundary prediction.

In summary, it appears we are very near optimal attainable performance in phrase boundary detection from Fo contours, with few possibilities of improvement by revisions in the computer program. For some ideas about further detailed improvements in the BOUND 3 phrase boundary detection program, see our previous semiannual report (Lea, 1976c). However, results with the 255 sentences suggest that the places where boundaries should be predicted could deserve further study.

In addition to detection of phrase boundaries from Fo contours, several other studies about Fo contours have been conducted at Sperry Univac. In 1973, Lea proposed a model of Fo contours in which effects due to clauses, phrasal groupings, stress patterns, and phonetic effects were superimposed. Our recent studies of Fo contours in the 255 all-sonorant sentences (section 3.5 of this report) confirm the general rapid-rise, gradual-fall intonation of declarative, WH, and command clauses, and the terminal rise that sometimes but not always accompanies yes/no questions. The peak of the Fo contour in a clause was shown to occur during or just after the first stressed syllable in the clause. Stressed syllables were almost always exhibited by local Fo rises at or near their syllabic onsets, as predicted in Lea's model. The effects of phonetic sequencies on Fo contours are superimposed on the clause, phrase, and stress effects, and are removeable by use of all-sonorant sequences. As has been repeatedly verified, Fo dips slightly during voiced obstruents, and is initially high after unvoiced obstruents and then followed by a rapid fall.

An interesting sidelight to our studies of fundamental frequency contours in the 159 sentences (Lea, 1976c) was the observation that large fundamental frequency variations occurred before many stressed word-initial vowels. These were obviously the result of glottal stops. The glottal stop is often preceded by a local rise in fundamental frequency, which suggests an acoustic (hence, universal physiological) origin of rising tones that are often found to precede glottal stops in tone languages. After a glottal stop, a rapidly rising fundamental frequency, or other major perturbations of fundamental

frequency, may occur. Unvoicing may be apparent during the glottal stop. Obviously, fundamental frequency variations thus may be indicative of the occurrences of glottal stops, so that they may be distinguishable from oral stops.

In addition, the results with the 159 sentences strongly indicate that glottal stops are more likely to occur before stressed vowels than unstressed ones. If a glottal stop occurs, it very probably precedes a stressed vowel, and is often likely to be just after a major constituent boundary. The glottal stop is thus another potential cue to stress and constituent structure.

In another study of Fo contours, Dean Kloker (1976), showed that over 80% of the perceived phrase boundaries in spectrally inverted speech were detectable by an Fo model that automatically locates and describes the shape of Fo patterns throughout a sentence. He used the function $y(t) = at^b e^{ct}$ to model the rise-fall shapes which define phrases, with parameters a, b, and c derived from a stepwise regression which adds new values to the region of a phrasal contour as long as the variance of the fit is not too large and the next Fo value is within a prediction interval. Twenty one percent of the boundaries found by the model were not found perceptually (i.e., were "false alarms"). Kloker also found that sentence-final phrases marked as <u>complete</u> clause boundaries by listeners were generally found to be falling or level patterns, while those heard as <u>incomplete</u> were all found to be rising Fo contours.

Our study of Fo contours in the contrasting structures of the "flying planes paradigm" (page 36) showed a clear case of how Fo contours, and the boundaries detected from them, can be used to distinguish between alternative syntactic bracketings of a sentence. No boundary occurs between the adjective and final noun in the NP-COPULATIVE-NP structure, while a boundary does occur between the verb and final noun of the NP-AUX-V-N structure. Many other structural contrasts can be possible where boundary detections can distinguish among alternative structural hypotheses.

As boundary locations are more precisely predictable and we know more specifically just which constituents will be accompanied by boundaries,

boundary detections can be used ever more effectively in hypothesizing syntactic structures given the prosodic information. Then, such prosodic adjustments of parsing paths as was attempted for the BBN HWIM system will be possible and desirable.

4.5 Perceived Stress Patterns

Before we could evaluate any automatic procedure for locating stressed syllables, we needed a "standard" specifying which syllables in speech are actually stressed. In an initial study in 1973, we had three listeners individually listen to portions of speech tapes, rewinding at will and listening again until they could mark each syllable as either stressed, unstressed, or reduced. Texts studied were the Rainbow Script spoken by six talkers, the Monosyllabic Script spoken by two talkers, and 13 of the ARPA man-computer interaction sentences (involving eight talkers). Each listener repeated the perception test three times, with trials separated by several days. With three repetitions with speech, three without speech (using only the written text), three listeners, and with the various speakers involved, this study involved a total of about 28,000 judgments of stress levels for syllables in the connected texts.

As expected, the different listeners sometimes assigned different stress levels to the same syllables, presumably based on how they individually defined the boundaries between categories of stressed, unstressed, and reduced syllables. Their confusions were not seriously increased or decreased in going from individual talker to talker, or from text to text. Two listeners were found to agree in their perceived stress levels for most of the individual syllables. They differed on only about 5% of all syllables as to whether they were stressed or not, and each of them showed only about 5% confusions in decisions about stressed syllables from one trial to another. Unstressed and reduced levels were much more frequently confused. A third listener differed from the other two listeners on about half of his stress level judgments, and also labelled substantial percentages of all syllables as stressed on one trial and unstressed on another. Such listeners who are inconsistent in their own judgments and who differ dramatically from other listeners should be excluded in any attempts to establish standards about which are the actual "stressed syllables" in connected speech.

The listeners also appeared to be as consistent in their assignments of stress levels given only the written text as they were in their assignments when listening to the speech recordings. However, their judgments without speech did not correspond well with their judgments with speech if the speech was spontaneous (that is, not produced by speakers reading written texts). Listeners appeared to differ most dramatically from each other, and yield more confusion in stress levels from repetition to repetition, when yes-no questions were involved. (Later studies with more questions did not show such a difference due to sentence type; Lea, 1976c.)

These initial studies were later extended to all the 31 ARPA sentences (Lea, Medress, and Skinner, 1973b) with similar results. Then, in 1976, with the completion of the design and recording of the large 3300-sentence speech data base, further studies were conducted on listener's perceptions of stress patterns, both for enhancing our understanding of the method of obtaining perceptions, and to supply stress judgments on a variety of sentence structures. These stress judgments provide the 'standard' of correct stress assignment by which acoustic correlates of stress can be evaluated, and also provide considerable evidence (involving over 17,000 perceptions) about the stress levels accompanying various word categories, and the effects of syntactic processes (e.g., subordination and coordination) on stress patterns.

By an initial experiment in which eleven listeners provided stress perceptions on three separate trials (spaced one week apart), we demonstrated that five new listeners that were substantially untrained about prosodic structures could successfully (i.e., consistently, and with agreement among listeners) categorize all syllables in connected speech as either stressed, unstressed, or reduced. Good listeners may be selected on the basis of consistency from time to time and agreement with other listeners. Listeners agree that, with a few times of rewinding and listening to the clauses in a sentence, they can effectively and meaningfully mark stress patterns. They usually listen first for stressed syllables throughout clauses, then fill in decisions about reduced and unstressed syllables.

It appears from these studies that the relative stressedness of syllables can be reliably determined from counting the number of listeners that agree that a syllable is stressed (or reduced), thus yielding a "stress score" which is highest for the most stressed syllables and lowest for the most reduced syllables. Using such a stress score, we have demonstrated that WH-words, nouns, quantifiers, and command verbs are among the most-stressed words in English sentences, and that main verbs, adverbs, adjectives, and negatives are also usually stressed. Auxiliary verbs, copulatives, pronouns, relative pronouns, and possessive determiners are usually unstressed, while articles, prepositions, and conjunctions are usually reduced. Coordination produces significant reductions in stress levels on repeated parts (verbs or nouns), and subordination of one clause or phrase under another causes reduction in stress scores on verbs, auxiliary verbs, and conjunctions, but not nouns.

Another structural regularity that was observed was that perceived stresses tend to decrease throughout a word sequence that would otherwise be expected to have equal stresses, or a rising stress pattern; stresses on subject nouns are higher than on direct object nouns, and stresses on prenominal modifiers (adverbs, adjectives, and participles) show a descending stress pattern, not the expected nuclear stress pattern.

In addition to providing extensive experimental evidence about English stress patterns, these studies have provided the necessary standard of correct stress assignment by which acoustic correlates of stress can be evaluated.

4.6 Automatic Location of Stresses

In 1972, we first proposed a strategy for locating stressed syllables. Based on previous studies that had shown that local increases in Fo and large integrals of energy within a syllable are the most reliable acoustic correlates of stress, this algorithm looked for regions of high energy integral near local Fo increases. The increasing Fo near the beginning of each constituent detected by the boundary detector was assumed to be attributable to the first stressed syllable in the constituent (Lea, 1973b, section 5). A stressed "HEAD" to the constituent was thus associated with a portion of the speech which is high in energy with rising Fo, and bounded by substantial (5 dB or more) dips in energy. Other stressed syllables in the constituent were expected to be accompanied by local increases in Fo. Since

the usual ("archetype") shape of the Fo contour in a constituent is a rapid rise followed by a gradual fall in Fo, we expected that local 'increases' in Fo due to later stressed syllables would show local rises above the gradually falling Fo contour, even if Fo did not rise absolutely near the stressed syllable. The stressed syllable is located within a high-energy-integral region near this local rise above the archetype Fo contour.

This strategy was first specified precisely and used in rigorous hand analyses but not implemented as a computer program until 1975. In the interim, the algorithm was tested by hand analysis of Fo and energy contours for 400 seconds of connected speech; namely, the Rainbow Script spoken by six talkers, the Monosyllabic Script spoken by two talkers, and, later, the 31 ARPA test sentences. The algorithm succeeded in locating an overall average of around 85% of all syllables perceived as stressed by the majority votes of a panel of listeners. Performance was best with speech read from written texts, but even in the 31 ARPA man-computer interaction sentences, over 85% of the perceived stresses were found. About 20% of all algorithmically located "stresses" were false, in that they did not point to syllables perceived as stressed by a majority of the listeners.

It was conceivable that simpler procedures for stress location might work as well as the archetype contour algorithm we had developed. Consequently, in 1973, Lea (1973f) did a comparison of three approaches to stressed syllable location. Methods based on only the durations of high energy chunks, or upon only the length of time that fundamental frequency (Fo) was not falling significantly, did not perform as well as the original algorithm based on archetype Fo contours in phrases and local searches for high-energy chunks of speech. The archetype contour algorithm was also least sensitive to the type of sentence being processed, while the other algorithms showed quite different performance in yes/no questions.

In March, 1975, the archetype algorithm was implemented as a FORTRAN program ("STRESS") and distributed to SUR contractors. This represented a major milestone in Sperry Univac's efforts to provide prosodic aids to speech understanding. The implementation included a number of improvements and new tests not included in the original algorithm, including refined methods for selecting the highest-energy nucleus near a rise in Fo, a method of picking up some stresses that were missed in very long phrases, and a special test for prepausal stresses. The program was tested with the Rainbow and Monosyllabic

Scripts, spoken by two talkers, and the 31 ARPA sentences. On the average, 89% of the syllables perceived as stressed were found by the program, while about one out of five locations were false. The STRESS program confused about 15% of all syllables between the stressed and unstressed categories, while listeners confused about 5% of the syllables. While the program is obviously open to some improvements, it is approaching the level of performance that listeners can attain.

In our most recent study of syllabification and automatic stress location (see section 3.4), we found that 91% of all the syllables in the 255 database sentences were correctly detected by the syllabification routine, while some weak syllables in all-sonorant sequences were missed. These results were surprisingly good, considering the difficulty of syllabification in all-sonorant sequences. The overall scores of 92% correct stressed syllable location and 20% false alarms were also very satisfying. There is, however, still room for considerable improvement. One such improvement is expected to come from a revision of the STRESS program to associate the first stress in a sentence with the nucleus immediately preceding the peak Fo value in the sentence. Another improvement would be a revision of the utterance-final (prepausal) test for stresses. Other possible improvements were discussed in section 3.4.

Steps should be taken in future studies to use the located stresses to guide phonemic analysis, word matching, and syntactic parsing processes.

4.7 Timing Cues to Linguistic Structures

In 1974, we conducted a study of rhythm and timing cues in our available speech texts (Lea, 1974a). These studies suggested that stressed syllables tend to occur at intervals of about 0.4 to 0.5 seconds (that is, there is some tendency toward stress "isochrony"). However, the variation in interstress interval sizes was quite large, even for a single talker within a single text. We concluded that the concept of English being a stress-timed language is not simply exhibited by exact equality of interstress intervals, or even by an unquestionable "tendency toward equality" of interstress intervals regardless of other factors. We found that, contrary to several published hypotheses, the average interstress interval increases about linearly with the number of unstressed syllables between the stresses. A tendency toward stressed-unstressed alternation was exhibited, and it is probably this tendency, plus the somewhat uniform durations expected for unstressed syllables, that yields the tendency for interstress intervals to cluster somewhat near an average of 0.4 seconds or so.

This study also showed that pauses between clauses of a sentence tended to be about the same duration as interstress intervals, while pauses between sentences tended to be twice that duration. A pause is thus like an integer multiple of an inserted silent interstress interval. We also found that time intervals between detected syntactic boundaries tended to cluster in a multimodal distribution centered around multiples of the average interstress interval.

These results indicate that interstress intervals, pause durations, and intervals between detected boundaries all seem to relate to speech rhythm. Each of these, plus a measure like the number of syllables per second, may be useful as a measure of the rate of speech. Information about rate of speech may be used in selecting the appropriate phonological rules to apply in determining underlying phonemic structure from the slurred, coarticulated phonetic sequences. "Fast speech" rules show more slurring, coarticulating, and dropping of speech sounds.

We experimentally investigated how various measures of the rate of speech correspond with changes in phonological structure that should be handled by "fast speech" phonological and acoustic phonetic rules. The duration of the interstress interval was found to inversely correlate with the percentage of phones that were erroneously categorized by various available methods for automatic phonetic categorization. Other measures of speech rate, such as the number of syllables per unit time, were not as closely correlated with phonetic error rates. The interstress interval thus appears to be useful in predicting phonological rules that might apply to an utterance.

In addition to such phonological use of rate of speech, specific rhythmic effects such as interruptions of rhythm (pauses, "disjunctures", etc.) could be useful in hypothesizing the grammatical structure of a sentence.

Two experiments were conducted that verified the occurrence of timing cues to grammatical structure. In one experiment, five sentences per speaker were selected from the speech of six individuals who participated in simulations of computer interactions. The utterances were distorted by spectral inversion and presented to five listeners who marked stressed syllables, and the locations and types (normal or hesitation) of phonological phrase boundaries, using only the prosodic cues remaining in the signal. Vowel and sonorant durations (with and without aspiration) were measured from spectrograms, and then declared

stressed or unstressed based on the perceptions. Exploring the hypothesis that large increases in phonetic duration are syntactically determined, perceived boundary locations were compared with preceding segments which were 20% above the median length for that segment type. Using a rule which groups lengthened syllables, and from the lengthened group predicts phrase boundaries, 91% of the perceived boundaries were predicted. Of all the perceived phrase boundaries, those before silences longer than 200 milliseconds were more reliably predicted by lengthening than boundaries not at long silences. Locations perceived to be normal phonological phrase boundaries were more reliably predicted than those perceived as hesitations. Of the predicted boundary locations not perceived by listeners, some marked major syntactic boundaries, but most were at minor syntactic breaks, notably between modifiers and nouns, and after prepositions. The results also suggested that speaker differences and style variations may be important.

In another experiment, the question was whether or not one could detect major phrase boundaries from timing of prosodic features alone (such as onsets of syllabic nuclei found from energy contours), without the need for a prior determination of the phonetic sequence or the detection of lengthening of phonetic segments. We have already noted that syntactically-dictated pauses appeared as one-or two-unit interruptions of rhythm. Interstress intervals spanning those pauses were thus two or three times their average duration within clauses. In addition long disjunctures (i.e., interstress intervals greater than 0.5 seconds) accompanied 95% of the perceived boundaries between phonological phrases. We thus do have quite reliable cues to linguistic structure in the timing of speech events.

4.8 Carefully Designed Speech Databases

From our initial "natural experiments" with the Rainbow, Monosyllabic, and 31 ARPA speechtexts, we were able to get initial evidence suggesting specific relationships between various acoustic prosodic features, on the one hand, and linguistic structures, perceptions, and abstract notions (such as boundaries, phrase structures, and rhythm, etc.), on the other hand. However, one cannot be certain with such natural experiments that some unknown third variable is not the source of any apparent relationships between the acoustic variable and the uncontrolled underlying abstract variable. Controlled experiments, with all variables except one fixed in the comparison of two

utterances, provide the proper extension from the encouraging results of the natural experiments. We consequently undertook the design of speech texts to provide the necessary controls and sufficient data to extend these encouraging tendencies into well-defined rules relating prosodic variables and linguistic structure. As noted in the previous sections, controlled tests with these sentences have contributed substantially to our understanding of prosodic structures.

There is a definite need to develop precise rules for systematically relating prosodic patterns to underlying structures. If one can understand how the interacting effects of semantics, syntax, lexical structures, stress patterns, and phonetic sequences are superimposed in the Fo and energy contours and time patterns of controlled English sentences, he has some of the most essential tools for using acoustic prosodic data to guide speech understanding strategies. Only by a systematic attack on the task of compiling experimentally-verified rules can one hope to provide the kind of reliability needed to make such prosodic data of major value in speech understanding systems. For example, our earlier predictions of where phrase boundaries should occur in Fo contours were based on intuitive analyses of syntactic structures. Where expected boundaries did not occur, or wherever false or unexpected boundaries occurred, there had been no recourse indicating the source of the error. This was in part due to the intuitive predictions used, and in part due to the uncontrolled syntactic structures involved in the texts studied. Experiments with the designed sentences of known syntactic structure have already begun to indicate exactly what structural boundaries are marked, and will ultimately permit the writing of precise rules predicting where boundaries will occur in new sentences of similar structures. These rules for predicting detectable boundaries may then be useful in computer determination of possible underlying structure given the detected boundaries. Similarly, precise rules for relating stress patterns to underlying structures are needed.

Of primary importance in such prosodic studies is the development of English intonation rules. Intonation rules cut across the whole gamut of problems involved in speech understanding, including the explanation of why constituent boundaries are detectable in Fo contours, what are the acoustic correlates of stressed syllables, how syntactic and semantic structures might be manifested

acoustically, what are useful phonological rules and morphological rules in various contexts, and how stress rules might be inferred from acoustic data. Many rules and hypotheses about regular prosodic patterns have been published, but few have been tested with extensive speech data. We consequently designed an extensive set of 922 sentences which provide "minimal pairs" of sentences with nearly identical word sequences but contrasting structures. These sentences include explicit tests of the prosodic effects of sentence type, contrastive syntactic bracketing, subordination, coordination, syntactic categories (such as pronouns, verbals, compound nouns, etc.), movement of stress within phrases, coreference, etc. Prosodic patterns that can be studied with these sentences include: performance of the program for detecting phrase boundaries from valleys in Fo contours; acoustic correlates of stressed syllables, and performance in automatic stressed syllable location; acoustic measures of rhythm and rate of speech; overall Fo contour shapes; and local variations in prosodic features due to phonetic sequences. Also designed was a set of 178 sentences which included all word-initial consonant-vowel (CV) sequences and all word-final vowel-consonant (VC) sequences. These "phonetic-sequence sentences" provide the speech data needed for efficiently testing automatic procedures for vowel and consonant classification. For example, five sentences provide instances of all distinguishable stressed vowels of American English, coupled with the sibilants (s,ʃ), in initial CV and final VC positions.

The 1100 designed sentences were recorded in a pseudorandom order by three male talkers, using unusual recording procedures that involved projecting the sentences one at a time on the wall of a sound proof room in which the talker was situated. Complete dialect information was obtained for the three talkers. Subsets of the sentences (including 99 by one talker, 37 by another, and 255 by the third) were dubbed into a useful order for subsequent prosodic analysis.

From extensive studies with such designed sentences, one could hopefully develop experimentally-validated intonation rules and other prosodic rules. These rules would then be used to guide parsing, semantic analysis, phonological analyses, and word matching procedures in future speech understanding systems.

Only a modest beginning on such studies has been completed within our ARPA program, but the data is available for further studies. In addition, our experience with the unusually monotonic speech of two of the talkers suggests that the sentences ought to be recorded by other talkers.

A report about the designed sentences and many prosodic hypotheses that they can be used to test has just been published and should be of service in any future studies with the large speech database. I would like to reiterate here that the design of such an extensive set of sentences with minimally-distinguished sentence structures is a very valuable result. To design and record such large volumes of speech, devise hypotheses to test, and arrange the data into subsets for analysis is a major task which spanned almost three years at Sperry Univac. Other researchers might avoid duplicate effort by adopting some of the sentence structures, and perhaps even the speech recordings, for their studies.

5. CONCLUSIONS AND FURTHER STUDIES

Table VII (pages 41 to 43) lists the many specific accomplishments of Sperry Univac's work on Prosodic Aids to Speech Recognition for ARPA. We need not summarize such specifics again here. Rather, we shall consider the general conclusions from this work, and final suggestions for further studies.

5.1 Conclusions

Prosodic information can and should be used to provide: reliable anchor points for efficient and accurate phonetic analysis; guidelines for phonological rule applications; segmentation of sentences into phrases; indications of syntactic features like sentence type, subordination, and coordination; and cues to semantic relations. We have shown not only why such use of prosodics is important, but, to some degree, how prosodics can be incorporated into speech understanding systems. Specific computer programs are now available for obtaining Fo contours, locating syllabic nuclei and syllable boundaries, determining which syllables are stressed, and segmenting speech into phrases.

A general strategy for prosodically guided speech understanding would: segment speech into phrases; locate the stressed syllables; do a phonetic analysis anchored around the reliable stressed syllables and other islands of phonetic reliability; hypothesize words that match that phonetic structure; postulate syntactic structures that match the prosodic patterns of phrase boundaries, stress patterns, timing, and intonation; hypothesize phrases or word sequences that match the prosodic, segmental, and lexical information; and verify semantic and pragmatic conditions. No such system has been developed, and none of the ARPA systems come close to using prosodics to the degree we have recommended. However, our initial cooperative effort with BBN, to incorporate intonational phrase boundaries into the parsing procedures of the BBN HWIM system, was giving encouraging results as the ARPA/SUR program came to a close. Also, syllabification, Fo tracking, and even rudimentary aspects of phrase boundary detection were incorporated into the systems.

Five years ago almost no mention was made of the role of prosodics in speech understanding systems. Prosodics were not listed among the major knowledge sources or "levels" of system organization in the original report of the ARPA study group that defined the ARPA/SUR program. One would have to look long and in a variety of directions to hear or see even any "lip

service" given to the possibility that prosodics should play any role in speech recognition.[1] Indeed, prosodics was at an infant stage comparable to that which phonemic structures had in the word recognition efforts of the early 1950's. Prosodics had less acceptability in 1971 than syntax enjoyed in Lindren's 1965 survey of work on speech recognition. Now, prosodics have reached such a level of acceptability and attention that one rarely hears a general prediction of future work without a substantial (though not always a well-informed) acknowledgement of the need for a prosodic knowledge source in the system. In a review of one of my recent papers, the reviewer put the question of whether to use prosody as now a foregone conclusion, and considered that now it is a question of how to use it. To a long-term advocate of such a "weak-sister" in the array of speech processing tools this is a heartening accomplishment.

Sperry Univac's efforts in the ARPA program began as basic supportive research on prosodics and their relationships to speech understanding systems. Only at a late stage in the program did the pressures for successful systems and closer cooperation among contractors lead us into concerted efforts to incorporate our ideas and experimental results into the systems being developed by other ARPA contractors. In retrospect, if such practical application of prosodics within systems was to be accomplished primarily by us, the researchers on prosodics, rather than by the system builders, such an orientation should have been taken earlier. We didn't quite make our experiments be of practical application until shortly before the systems had to be frozen for performance evaluation.

Consequently, our biggest accomplishments were in the area of experimental studies of prosodic structures. This is evidenced by the long list of experimental results listed in Table VII (pages 41 to 43). We provided solid experimental evidence for what were intuitively accepted notions about the value of prosodics; namely, that any of various available methods of automatic labelling of phonetic segments worked best in the carefully articulated stressed syllables, that intonation provides cues to phrase structure, that stress patterns can be used to directly detect some aspects of syntactic

1 A notable exception was the work of the late Gordon Peterson (1961, 1963) perhaps the earliest spokesman for the use of prosodics in speech recognition.

structure, and that stresses are crucial to the rhythm, rate, and prediction of phonological distortions in speech.

We experimentally confirmed or disproved linguistis' and theorists' rules about expected stress patterns, intonation contours, pauses, rhythms, and perceptions of prosodics. Some of our results have major theoretical significance to linguists and speech scientists. For example, contrary to Bolingers' published claims (1965, 1972), there is a neutral, syntactically-determined intonation contour and stress pattern for spoken English sentences. Contrary to the claims of Pike (1945) and other linguists, isochrony of English stresses is not exhibited by simple squeezing of unstresses between fairly fixed onset times of stressed syllables. Some aspects of claimed "nuclear stress patterns" (cf. Chomsky and Halle, 1968) are not confirmed by either perceived stress patterns or acoustic correlates. Despite criticisms (Armstrong and Ward, 1929; Lieberman, 1967) of the close association assumed between constituent structures and invariant prosodic signals, there is considerable evidence that major syntactic boundaries are reliably marked by pauses, Fo contours, phrase-final lengthening of vowels and sonorants, longer interstress intervals, and even specific phonetic segments like glottal stops. A major challenge to total language models is the distinction between the positions of surface syntactic boundaries and the displaced indicators of those boundaries in Fo contours and groups of lengthened syllables. Similarly, though the subject-predicate boundary is considered among the major syntactic breaks in a sentence (cf. e.g. Scholes, 1971), that boundary is one of the least detectable from the prosodic patterns we have studied.

Our research has spanned the whole gamut of prosodic structures, and I believe it provides vital background for further work on prosodic aids to speech recognition. We are very close to where prosodics can be used to provide valuable aids to phonological analysis, word matching, and parsing. Indeed, if a system builder cannot now accept the need for a total prosodically guided speech understanding strategy such as we have proposed, he should at least give careful consideration to incorporating a "prosodic verifier" which compares expected prosodic patterns for hypothesized word sequences with actual detected patterns, and thus adjusts scores of alternative hypotheses. Such ideas are being explored by Sperry Univac under internal funding.

Many of our most solid results about prosodic structures have come from the recent use of our database of sentences with minimal pairs of contrasting structures. Such designed sentences should play a valuable role in any prosodics research or development of prosodic aids to speech understanding.

5.2 Further Studies

If I had my unrestricted choice and the necessary resources to undertake a program in prosodic aids to speech recognition today, I would do the following, and I obviously recommend this approach to interested researchers. I would have a two-prong effort: (1) conducting necessary experimental research on prosodic regularities; and (2) developing a predominently prosodically-guided speech understanding system which can progressively incorporate more prosodic information.

We do not know all we need to know to be able to simply apply prosodics to speech understanding without simultaneous further research. Anyone who would promote prosodic aids without further experimentation would be in danger of slowing the ultimate progress of prosodically guided systems by a premature application of limited information. Such a tactic may even lead to discouragement about prosodics, resulting from ill-devised and improperly applied limited tools. We need to know more precisely: just which constituents are demarcated by Fo contours and other prosodic cues; what intonation can tell us about sentence type, subordination, coordination, and special phrase structures, how to remove or handle phonetic influences on prosodics; what stress patterns can actually be expected with various phrase structures; how to use timing cues to select phonological rules; etc. We also need to test prosodic regularities with more talkers, other speech styles, and more repetitions per talker. In general, useful rules for relating prosodic patterns to linguistic structures must be experimentally developed.

On the other hand, the development of useful rules for relating prosodic patterns to linguistic structure also demands the direct application of those rules to working systems, to evaluate their accuracy and utility. I would recommend a system structure that makes use of prosodics from the very beginning of system implementation. Two alternative beginnings are (A) a prosodically guided speech understanding system such as we have previously defined (Lea, 1974; Lea, Medress & Skinner, 1975); or (B) a more standard system with acoustic analysis, phonetic segmentation, word matching and scoring, and appropriate parsing and control structures that hypothesize and test word sequences, but with a "prosodic verifier". The prosodic verifier would

compare expected stress patterns with detected stress patterns to adjust word scores, compare expected and Fo-detected phrase boundaries to adjust scores on word hypotheses, compare prosodic indicators of sentence type with the hypothesized type of sentence, etc. Either system could use stressed syllables as phonetically-reliable anchors around which a search for occurrences of words can be attempted, and, if convenient, they could restrict expensive acoustic analyses such as LPC spectral analysis to only those regions (voiced regions or maybe only stressed syllables) where prosodics could suggest that such analysis is needed.

The system should be used initially for very restricted tasks with few syntactic structures, comparable to or more restricted than those used in the successful HARPY and HEARSAY systems. Later work could deal with more challenging tasks such as the versatile subset of English handled within the BBN HWIM system.

We need to define precise ways of using prosodics in word matching and parsing. Can one reliably rule out words from hypothesized occurrence at a certain point in an utterance, based on the wrong syllables being stressed or phrase boundaries occurring where they are not expected? Can one rule out (or reduce the score on) possible phrase structures because the phrase boundaries that were detected are at radically different places from those predicted for those structures? Even for very restricted speech recognition systems these ideas would seem worth incorporating and testing.

On the experimental side, I would continue testing the BOUND3 Fo-boundary detector, the syllabification routine, and the STRESS stressed syllable locator, using the remainder of our 922 "Prososyntactic Sentences" spoken by talker WAL, and, very soon, introduce other talkers for the same sentences. Later I would introduce repetitions of the same sentence by the same talker. Other syntactically and prosodically informative sentences might then be added. After some concrete results with several talkers, I would explore similar questions with other speech styles.

Early attention should be directed toward the following questions: Which constituents are marked by prosodic boundaries?; What is the success in boundary detection, syllabification, and stressed syllable location?; What regularities are to be found in initial, terminal, and medial Fo contours

within clauses? Can one find more adequate procedures for extracting syntactic and stress-related aspects from phonetic influences on Fo contours? From the beginning, one should see a primary goal of developing experimentally-verified intonation rules. About the time that more talkers are introduced, I would recommend thorough studies of all acoustic correlates of stress, with the thought of improving or replacing the current stress location program. Studies of rhythm, rate, and the use of interstress intervals to predict applicable phonological rules should be undertaken by the time that significant data is available from several talkers.

In essence, I am saying that we are in the middle of the necessary experimental research about prosodic structures, with considerable work yet to be done, but with the possibility of promptly beginning to apply restricted prosodic information within a speech understanding system. While the ARPA/SUR program is over, and with it the excellent interactions, cooperative spirit, and interchange of ideas that has so much permeated that program, the need for speech understanding systems and for prosodic guidelines will not diminish, but rather increase. Speech understanding remains one of the most challenging potential users of prosodics, though, as I have noted previously (Lea, 1976c, pp. 49-50), prosodics can also be used in other systems for word and concept spotting, language identification, speaker identification, and speech synthesis.

6. REFERENCES

ALLEN, J. and O"SHAUNESSY, D. (1975). "Fundamental Frequency Contours of Auxiliary Phrases in English," MIT/RLE Technical Report, Massachusetts Institute of Technology, Cambridge.

ANDERSON, B. F. (1966). The Psychology Experiment. Belmont, California, Brooks/Cole Publishing Company.

ARMSTRONG, L.E. and WARD, I.C. (1926). Handbook of English Intonation. Cambridge; Heffer (2nd Edit.).

BOLINGER, D. L. (1965). Forms of English: Accent, Morpheme, Order. Cambridge: Harvard Univ. Press.

BOLINGER, D. L. (1972). "Accent is Predictable (if Your're a Mind Reader)," Language, vol. 48, 633-644.

CHEUNG, J. Y. (1975). "Computer Estimates and Modelling of Linguistic Stress Patterns in Speech", Ph.D. Dissertation (EE Tech. Report No. 188), Dept. of Electrical Engineering, Univ. of Washington, Seattle.

CHOMSKY, N. and HALLE, M. (1968), The Sound Pattern of English. New York: Harper and Row.

KLOKER, D. R. (1975). "Vowel and Sonorant Lengthening as Cues to Phonological Phrase Boundaries," presented at the 89th Meeting, Acoustical Society of America, Austin, Texas, April 8-11. Abstract in J. Acoust. Soc. America, vol. 57, Supp.

LEA, W. A. (1972). "Intonational Cues to the Constituent Structure and Phonemics of Spoken English," Ph.D. Dissertation, School of Electrical Engineering, Purdue University.

LEA, W. A. (1973a). "Syntactic Boundaries and Stress Patterns in Spoken English Texts," Univac Report No. PX 10146, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1973b). "Segmental and Suprasegmental Influences on Fundamental Frequency Contours." In Consonant Types and Tone (Proceedings of the First Annual Southern California Round Table in Linguistics, Ed. by L. Hyman). Los Angeles: University of Southern California Press.

LEA, W. A. (1973c). "An Approach to Syntactic Recognition Without Phonemics", IEEE Trans. on Audio and Electroacoustics, vol. AU-21, 249-358.

LEA, W. A. (1973d). "Evidence that Stressed Syllables Are the Most Readily Decoded Portions of Continuous Speech," Paper presented at the 86th Meeting of the Acoustical Society of America, Los Angeles, California.

LEA, W. A. (2973e. "Perceived Stress as the 'Standard' for Judging Acoustical Correlates of Stress," Paper presented at the 86th Meeting of the Acoustical Society of America, Los Angeles, California.

LEA, W. A. (1973f). "An Algorithm for Locating Stressed Syllables in Continuous Speech," Paper presented at the 86th Meeting of the Acoustical Society of America, Los Angeles, California.

LEA, W. A. (1974a). "Prosodic Aids to Speech Recognition: IV. A general Strategy for Prosodically-Tuided Speech Understanding," Univac Report No. PX 10791, Sperry Univac, DSD, St. Paul, Minnesota.

LEA, W. A. (1974b). "Sentences for Controlled Testing of Acoustic Phonetic Components of Speech Understanding Systems," Univac Report No. PX 10952, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1975a). "Isochrony and Disjuncture as Aids to Syntactic and Phonological Analysis," presented at the 89th Meeting, ASA, Austin, Texas.

LEA, W. A. (1975b). "Prosodic Aids to Speech Recognition: VII. Experiments on Detecting and Locating Phrase Boundaries," Univac Report No. PX 11534, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1976a). "Acoustic Correlates of Stress and Juncture," Univac Report No. PX 11693, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1976b). "The Importance of Prosodic Analysis in Speech Understanding Systems," Univac Report No. PX 11694, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1976c). "Prosodic Aids to Speech Recognition: VIII. Listeners' Perceptions of Selected English Stress Patterns, Univac Report No. PX 11711, Sperry DSD, St. Paul, Minnesota.

LEA, W. A. (1976d). "Use of Intonational Phrase Boundaries to Select Syntactic Hypotheses in a Speech Understanding," presented at 92nd Meeting, Acoustical Society of America, San Diego. J. Acoust. Soc. America, vol. 60, Supplement 1, Fall 1976, S12(A).

LEA, W. A. (1976e). "Sentences for Testing Prosodic and Syntactic Components of Speech Understanding Systems," Univac Report No. PX 10953, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (1972a). "Prosodic Aids to Speech Recognition I: Basic Algorithms and Stress Studies," Univac Report No. PX 7940, Univac Park, St. Paul, Minnesota.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (1972b). "Use of Syntactic Segmentation and Stressed Syllable Location in Phonemic Recognition." Presented at the 84th Meeting, Acoustical Society of America, Miami Beach, Florida, Nov. 27-30, 1972.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (1973a). "Prosodic Aids to Speech Recognition: II. Syntactic Segmentation and Stressed Syllable Location," Univac Report No. PX 10232, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (1973b). "Prosodic Aids to Speech Recognition:III. Relationships between Stress and Phonemic Recognition Results," Univac Report No. PX 10430, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. MEDRESS, M. F., and SKINNER, T. E. (1975). "A Prosodically-Guided Speech Understanding Strategy," IEEE Trans. Acoustics, Speech and Signal Processing, vol. ASSP 23, 30-38.

LIEBERMAN, P. (1967). Intonation, Perception, and Language. Cambridge: M.I.T. Press.

MEDRESS, M. F. (1972). "Revised Computer Phonetic Transcriptions," ARPA/SUR unpublished SUR Note No. 32, Sperry Univac DSD, St. Paul, Minnesota, April, 1972.

OLLER, D. K. (1973). "The Effect of Position in Utterance on Speech Segment Duration in English," J. Acoust. Soc. America, 1235-1247.

O"SHAUGHNESSY, D. (1976). "Fundamental Frequency by Rule for a Text-to-Speech System," presented at the 92nd Meeting, Acoust. Soc. of America, San Diego. J. Acoust. Soc. of America, vol. 60, Suppl. No. 1, S76(A).

SARGENT, D. C. (1975). "Computer Algorithms for the Extraction and Application of Stress Contours from Continuous Speech Sentences," Report No. TR-EE 75-44, School of Electrical Engineering, Purdue University, West Lafayette, Indiana, 47907.

SCHOLES, R. J. (1971). Acoustic Cues for Constituent Structures. The Hague: Mouton and Co.

SKINNER, T. E. (1973a). "Speech Parameter Extraction: Fundamental Frequency, Spectral and Formant Frequency Processing," Univac Report No. PX 10376, Univac Park, St. Paul, Minnesota.

7. APPENDICES

APPENDIX A. BOUNDARIES AND STRESS PATTERNS IN THE 255 DATABASE SENTENCES

In the following pages of Appendix, the 255 sentences used in our recent studies of acoustic prosodic patterns are listed. The sentences are grouped into subsets which test specific syntactic or lexical effects on prosodic structures. Each sentence is preceded by an identifier consisting of: (a) the letters "PSS" meaning "prososyntactic sentence", in contrast to phonetic sentences, etc.; (b) either S (short) M (medium), L (Long), or X (referring to phonetic or extra structural tests), along with a number identifying the sentence's place in the ordered description of the database; (c) a prediction of the stress pattern (stressed or not for each syllable), with parantheses around phrases; and (d) a tree number (e.g., T4), indicating the syntactic tree that represents the surface structure of the sentence. These identifiers are described more fully in another report (Lea, 1976e).

Also accompnaying the sentences are markings of the perceived and automatically detected prosodic patterns. Above each syllable is a number between -5 and +5) specifying the stress score (SS) for that syllable, where -5 means all five listeners heard the syllable as reduced, while +5 indicates all heard it as stressed. (See Lea, 1976c). Each syllable (or portion of speech including more than one syllable) which was automatically located as a stressed syllable is underlined. Thus, only syllables with stress scores of +3, +4, or +5 should end out being underlined. Any underlined portion that does not include a syllable perceived as stressed (that is, any portion with no SS ≥ 3) is a false alarm in stress location. Any syllable with SS $\geq$ +3 that is <u>not</u> underlined is a missed stress.

Another form of information displayed on the sentences concerns phrase boundaries detected from Fo contours. Every detected phrase boundary is marked by a vertical bar approximately at the position in the utterance where it was detected. Each position where a boundary was expected but not detected (i.e., a missing boundary) is shown by a star.

SUBSET 1A. One Stress Per Constituent

PSS S2-SS,T1
+5 +4
Men* know.

PSS S4-SUS,T2
+5 +1 +4
|Men will kn|ow.

PSS S8-SSS,T3
+5 +2 +3
Men* know* Ron.

PSS S12-SUSS,T4
+5 -1 +4 +2
Men will | know* Ron.

PSS S16-SSSS,T5
+5 +4 +5 +2
Men* know | Ron* now.

PSS S24-SSSS,T6
+5 +3 +5 +5
Men* owe* Ron | rum.

PSS S26-S(US),T1
+5 -1 +4
Men en* roll.

PSS S28-S(SU),T1
+5 +5 -2
Men | worry.

PSS S37-SU(US),T2
+5 -2 -3 +5
Men will en* roll.

PSS S39-SU(SU),T2
+5 -3 +4 -1
Men will | worry.

PSS S42-SS(US),T3
+5 +5 -1 +4
Men* know | Marie.

PSS S42-SS(SU),T3
+5 +5 +4 -1
Men* know* Mary.

PSS S47-S(US)S,T3
+5 -3 +4 +4
Men* en roll* Ron.

PSS S51-S(SU)S,T3
+5 +5 -2 +5
Men | worry | Ron.

PSS S56-SS(UUSU),T3
+5 +3 +2-5+5-3
|Men* know | Leonora.

PSS S57-SS(USU),T3
+5 +3 -2 +5-3
Men* know Ma|ria.

PSS S58-SS(SUU),T3
+5 +5 +5 -5 -1
Men* know | Melanie.

SUBSET 1A. One Stress Per Constituent. (cont.)

PSS	S66-S(US)(US),T3	+5 -1 +4 -1 +5 |Men* en roll Ma |rie.
PSS	S67-S(US)(SU),T3	+5 -3 +3 +5 -2 Men en roll* Mary.
PSS	S68-S(SU)(US),T3	+5 +4-2 -2 +5 |Men |worry Ma |rie.
PSS	S69-S(SU)(SU),T3	+5 +4 -2 +4-2 Men |worry |Mary.
PSS	S107-S S(UUS),T3	+5 +4 -3 -2 +5 Ron* knew a ma |rine.
PSS	S108-SS(USU),T3	+5 +3 -3 +5 -2 |Ron * knew an |airman.
PSS	S136-S U(US)(SU),T4	+5 -2 -2 +4 +5 -2 Ron will en |roll |airmen.

SUBSET 1B. Two or More Stresses Per Constituent:
Expansions of Determiner.

```
                                   +5  -1   -1   +4     +5   +5
PSS  S137-SU(US)(SS),T4            Ron will en* roll|  nine men.

                                   +5  -1   -3  +4   -2    +5
PSS  S138-SU(US)(SS)*,T4           Ron will en|roll your* men.

                                   +5  -1   -3  +4    +5  +5
PSS  S139-SU(US)(SS),T4            Ron will en|roll| all men.

                                   +5  -1   -3  +4    +5  +5
PSS  S140-SU(US)(SS),T4            Ron will en|roll  |no men.

                                   +5  -1   -2  +4    +4  +5
PSS  S141-SU(US)(SUS),T4           Ron will en|roll  |any men.

                                   +5  -1   -3   +4    +5 -3  +5
PSS  S142-SU(US)(SUS),T4           Ron will en * roll| many  |men.

                                   +5  0    -2    +4    +5    +4 -2
PSS  S143-SU(US)(SUS),T4           Ron will en* roll  |nine| airmen.

                                   +5  0    -2  +4   -1     +5 -2
PSS  S144-SU(US)(SSU)*,T4          Ron will en |roll your | airmen.

                                   +3  0    -2  +2    +1  -3    +4
PSS  S145-SU(US)(SSS)*,T4          Ron will en |roll* all your | men.

                                   +5  -1   -3  +4    +5  +5   +4
PSS  S146-SU(US)(SSS),T4           Ron will en |roll* all nine men.

                                   +5  -1   -3  +4     +5-3 +5   +4
PSS  S147-SU(US)(SUS[illegible],T4 Ron will en |roll | any nine men.

                                   +5  -1   -3  +4    +5  +5   +5  -3
PSS  S148-SU(US)(SSU),T4           Ron will en|roll | all nine airmen.

                                   +5  0    -3  +5   +5-2  +5   +4 -3
PSS  S149-SU(US)(SUSSU),T4         Ron will en|roll | any nine airmen.
```

*The initial prediction was that "your" would be stressed. It now seems more likely that "your" will be unstressed in these sentences.

SUBSET 1C. Prenominal Adjectives, Participles, and Adverbs (with 4 or less syllables in NP)

```
                                 +5   -1   -4  +3     +4    +5
PSS  M1-SS(US)(SS),T4            Ron will en |roll| young men.

                                 +5   -1   -3   +2    +5 -3   +4
PSS  M3-SS(US)(SUS),T4           Ron will en |roll| moral | men.

                                 +5   -1   -5   +3    +5  -3  +5
PSS  M4-SS(US)(SUS),T4           Ron will en* roll | willing men.

                                 +5   -1   -5   +4    +5    +5 -3
PSS  M5-SS(US)(SSU),T4           Ron will en |roll| young airmen.

                                 +5   -1   -5   +3  -5    +5   +5
PSS  M6-SS(US)(USS),T4           Ron will en |roll a | young man.

                                 +5   -1   -5   +3    +5   +4    +4
PSS  M8-SS(US)(SSS),T4           Ron will en |roll| nine young men.

                                 +5   -1   -5   +5   -3    +2    +5
PSS  M9-SS(US)(SSS)*,T4          Ron will en |roll your |young men.

                                 +3   -1   -5   +4    +5   +4    +5
PSS  M10-SS(US)(SSS),T4          Ron will en |roll| mean young men.

                                 +5   -1   -5   +4    -3 +4-5   +5
PSS  M11-SS(US)(USUS),T4         Ron will en |roll| immoral| men.

                                 +5   -1   -5   +4  -5  +5 -4   +5
PSS  M12-SS(US)(USUS),T4         Ron will en |roll a | moral man.

                                 +5   +3   -2   +4  -5   +5     -1   +4
PSS  M13-SS(US)(USUS),T4         Ron will en |roll a |young| ma rine.

                                 +5   -1   -4   +4    +5 -3 -2   +5
PSS  M14-SS(US)(SUUS),T4         Ron will en |roll| mannerly | men.

                                 +5   -1   -4   +4  -5   +5   +5 -3
PSS  M15-SS(US)(USSU),T4         Ron will en |roll a |young airman.

                                 +5   -1   -4   +4    +5-2  +4    +5
PSS  M16-SS(US)(SUSS),T4         Ron will en |roll| any young men.

                                 +5   -1   -4   +4    +5 -2  +4    +5
PSS  M17-SS(US)(SUSS),T4         Ron will en |roll| many young men.

                                 +5   -1   -4   +4    +5 -2  +3    +4
PSS  M18-SS(US)(SSUS),T4         Ron will en* roll | only young men.
```

*The initial prediction was that "your" would be stressed. It now seems more likely that "your" will be unstressed in these sentences.

SUBSET 1C. Prenominal Adjectives, Participles, and Adverbs (with 4 or less syllables in NP) (Cont.)

```
                              +5  -1   -4  +4     +5   +5 -3  +4
PSS  M19-SS(US)(USSS),T4      Ron will en* roll  |nine moral men.

                              +5  -1   -4  +5 -5  +4   +3    +5
PSS  M20-SS(US)(USSS),T4      Ron will en|roll a| new young man.

                              +5  -1   -4  +5  -2  +5    +3    +5
PSS  M21-SS(US)(USSS),T4      Ron will en|roll a| mean young man.

                              +5   -1   -4  +4     +5   +3    +5 -2
PSS  M22-SS(US)(SSSU),T4      R|on will en|roll  |nine young airmen.

                              +5  -1   -4  +4   -2     +3    +5  -1
PSS  M23-SS(US)(SSSU)*,T4     Ron will en|roll your| young airmen.

                              +5  -1   -5  +4    +5  -1     +3     +5
PSS  M24-SS(US)(SSSS)*,T4     Ron will en|roll| all your| young  men.

                              +5  -1   -4  +4   -2     +4   +4    +5
PSS  M25-SS(US)(SSSS)*,T4     Ron will en|roll your| nine young men.

                              +5   -1   -5  +4   -2     +4  +3    +5
PSS  M26-SS(US)(SSSS)*,T4     R|on will en|roll your| new young men.

                              +5  -1   -5  +4    +4  +4   +2    +4
PSS  M27-SS(US)(SSSS),T4      Ron will en|roll* new mean young men.
```

---

*The initial prediction was that "your" would be stressed. It now seems more likely that "your" will be stressed in these sentences.

SUBSET 1D. Prenominal Adjectives, Participle, and Adverbs (with more than 4 syllables in NP)

```
                                 +5   -1   +4   -4  -4  +4 -5     +2   +2
PSS  M28-SU(US)(USUUS),T4        Ron will en|roll  a  |moral  | marine.

                                 +5   -1  -5   +4   -4 -1  +5 -3   +5
PSS  M29-SU(US)(UUSUS),T4        Ron will en|roll an  im|moral man.

                                 +5   -1  -5   +4   -3  +4 -3    +5  -3
PSS  M30-SU(US)(USUSU),T4        Ron will en|roll  im|moral|  airmen.

                                 +5   -1  -3   +4  -2   +4 -3   +5  -2
PSS  M31-SU(US)(USUSU),T4        Ron will en|roll a   |moral  airman.

                                 +5   -1  -5   +4  -4    +5  -1   +4  -2
PSS  M32-SU(US)(USUSU),T4        Ron will en|roll a   |lonely  airman.

                                 +5  -1   -5   +4  -4   +5 -4   +4     +3
PSS  M33-SU(US)(USUSS),T4        Ron will en|roll a   |moral  young  man.

                                 +5  -1   -5   +4  -4     +5   +3  -3   +4
PSS  M34-SU(US)(USSUS),T4        Ron will en|roll a   |young moral  man.

                                 +5  -1   -5   +4     +5 -1    +5  -1   +4
PSS  M35-SU(US)(SUSUS),T4        Ron will en|roll    |moral |  lonely  men.

                                 +5  -1   -5   +4     +5  -1    +4 -3   +4
PSS  M36-SU(US)(SUSUS),T4        Ron will en|roll  |  lonely | moral  men.

                                 +5  -1   -5   +4     +5 -3  +4 -3  +5
PSS  M37-SU(US)(SUSUS),T4        R|on will en|roll  | many moral men.

                                 +5  -1   -5   +4     +5 -1    +5 -3    +5
PSS  M38-SU(US)(SUSUS),T4        Ron will en* |roll | only |  moral  | men.

                                 +5  -1   -5   +4    +5 -4    +4   -2   +4
PSS  M39-SU(US)(USUUS),T4        Ron will en|roll | many | willing  men.

                                 +5  -1   -5   +4    +5   -4  +5 -4  +5
PSS  M40-SU(US)(SUSUS),T4        Ron will en|roll | nine  im|moral men.

                                 +5  -1   -5   +4    +4-3     +4   -4  +4
PSS  M41-SU(US)(SUSUS),T4        Ron will en|roll | any  | young marine.

                                 +5  -1   -5   +4    +5    +5 -3    +5  -2
PSS  M42-SU(US)(SSUSU),T4        Ron will en |roll* nine |moral  | airmen.

                                 +5  -1    -5   +4   +5   -1   +4    +3    +5
PSS  M43-SU(US)(SUSSS),T4        Ron |will en| roll |lonely mean young men.

                                 +5  -1   -5   +4    +5  +4  -4    +3   +5
PSS  M44-SU(US)(SSUSS),T4        Ron will en| roll | new moral young men.
```

SUBSET 1D. Prenominal Adjectives, Participle, and Adverbs (with more than 4 syllables in NP) (cont.)

```
PSS  M45-SU(US)(SSSUS),T4      +5   -1   -5   +4     +5   +5    +4 -4     +5
                               Ron will en* roll | new young moral | men.

PSS  M46-SU(US)(SSS),T4        +5  -1   -5  +4     +5    +3    +4
                               Ron will en|roll | well known men.

PSS  M47-SU(US)(SUSS),T4       +5  -1   -5  +4     +5 -4   +4    +4
                               Ron will en|roll | really young men.

PSS  M48-SU(US)(SUUSUS),T4     +5  -1   -5  +4     +5 -2   -2 +4 -4    +5
                               Ron will en|roll | really immoral | men.

PSS  M49-SU(US)(SSUUSS),T4     +5  -1   -5  +4     +5 -2     +5 -4-3
                               Ron will en|roll |really | mannerly
                                +3      +5
                               young | men.

PSS  M50-SU(US)(SUSSS),T4      +5  -1   -5  +4     +4 -3     +3   +2
                               Ron will en* roll | really |well known
                               +4
                               men.

PSS  M51-SU(US)(SUSUSUSS),T4   +5  -1   -5  +4     +4 -2    -4+4 -4
                               Ron will en|roll |really immoral|
                               +4   +3     +4
                               well known | men.

PSS  M52-SU(US)(SUSSUSUS),T4   +5  -1   -5 +4     +4  -2    +3   +2
                               Ron will en|roll |really| well known
                               -4 +4 -4    +5
                               im|moral | men.

PSS  M53-SU(US)(SUSUUSUS),T4    +5  -1   -5  +4     +5 -3   +5  -1
                               R|on will en* roll |really| willing
                               -4 +4 -3   +5
                               immoral | men.

PSS  M54-SU(US)(SUSUUSUSS),T4  +5  -1   -5 +4     +5 -3    +4  -2
                               Ron will en |roll | really | willing
                               -4 +5 -4  +3    +5
                               im|moral young men.
```

SUBSET 1E. "Flying-Planes" Paradigm

PSS M229-(US) U (SUS),T3
+5 +1 -1 +5 -1 +5
Lawmen are | lying men.

PSS M230-(SU) U (SU)S,T4
+5 +1 -2 +5 -1 +5
Lawmen are | ruling | Maine.

PSS M231-(US) U (SUS),T3
+5 -1 -1 +5 -2 +3
Airmen ar|e lying men.

PSS M232-(SU) U (SU)(SU),T4
+5 -3 -1 -4 -2 +5 -3
Airmen are | eyeing | women.

PSS M233-(SU) U (SUS),T3
+5 -2 -1 +5 -2 +4
Airmen ar|e erring men.

PSS M234-(SU) U (SU) S, T4
+5 -4 -1 +5 -2 +4
Women are | airing | wool.

PSS M235-(SU) U (SUS),T3
+5 -1 -2 +4 -1 +5
Lawmen are | rummy men.

PSS M236-(SU) U (SU) (SU),T4
+5 -1 -2 +4 -1 +5 -3
Lawmen are | ruling | women.

SUBSET 2C$_2$. Movement of stress in the first constituent.

PSS S103-(US) S S,T3
-4 +5 +5 +5
A man* knew* Ron.

PSS S104-(USU) SS,T3
-2 +5 -4 +3 +5
A woman | knew* Ron.

PSS S105-(USU) SS,T3
-4 +5 -4 +5 +5
An airman* knew* Ron.

PSS S106-(UUS) SS,T3
-3 0 +4 +5 +5
|A marine | knew* Ron.

PSS S109-(US) S (US),T3
-4 +4 +2 -4 +5
M |onroe* knew a* man.

PSS S110-(US) S (US),T3
-4 +5 +4 -2 +5
A man* knew Ma |rie.

PSS S111-(SU) S (US),T3
+5-1 +5 -3 +4
M |ary* knew a | man.

PSS S112-(USU)S(USU),T3
-4 +5-2 +4 -4 +4 -3
Maria | knew an | airman.

PSS S113-(USU) S(USU),T3
-3 +5 -3 +4 0 +5-3
A woman | knew*Ramona.

PSS S114-(USU) S (USU),T3
-2 +5 -3 +5 -3 +4 -2
An a|irman | knew Ra|mona.

PSS S115-(USU) U (US) (UUS),T3
-3 +5 -3 -1 -5 +4 +1 -4 +5
A|n airman will en|roll | Marianne.

PSS S116-(UUSU) U (US)(UUS),T3
+2 -4 +5-2 -1 -4 +4 -4 -3 +5
Le |onora | will en|roll a ma |rine.

PSS S117-(UUSUU)U(US)(UUS),T3
-4 -1 +5-4-3 -1 -5 +4 -4 -3 +4
An Armenian | will en |roll a m |arine.

PSS S118-(UUS)U(US)(UUSU),T3
-4 -4 +5 -1 -4 +4 +1-4+5 -2
A marine will en|roll | Leonora.

PSS S119-(UUS)U(US)(UUSUU),T3
-3 -5 +5 -1 -5 +4 -5 -2 +5 -4-2
|A marine will en|roll an Ar |menian.

SUBSET 3D. Verb/Noun in Pairs

```
PSS  X15-(SSSU)(USU)(USUU),T3       -2  +4   +5  0      +4  0   -5  -1
                                    Our new | object | increases in-

                                    +4 -3-5 -1
                                    accuracy.

                                     +5-1  +5  -2   +4    -2   +5   -1   -5
PSS  X16-(SUS)(US)(USUU)(USUU),T243 Very few ob|ject|  to | increas|es

                                     -2  +5 -3 -5-1
                                    |in a |ccuracy.

PSS  X17-(UUSU)(US)(US)(UUSU)(USU) U (SU)(USU), T244

                                     -4  -3  +5  -3  -2   +5  -1  +4
                                    The com|puter can|not permit|

                                    0-5  +3  -2    -4    +5  0    -3
                                    vio|lations of  |syntax  |or

                                    +4   -1     -4    +5  -2
                                    conflicts in | schedule.

                                          -5  +5  -3   -4  -5
PSS  X18-(USU)(UUSUSU)U S U (USU) U (USU),T245    The record | of a

                                     +5  +1    -3  +3   -1   +5  -2  -5
                                    firearm | per|mit will |show if the

                                     +5  -1    -4 -4   +4   0
                                    suspect | is a | convict.

                                       -5  +3   +5  -1    -1   +3  -5
PSS  X19-(USSU)(US)(USU)(USU), T246    The two records | per|mit a|

                                     +5   0    -4    +5   -4
                                    conflict | in | schedule.

                                       -2   +5 -1     -3  +5   -2
PSS  X20-(USU)(US)(UUSU)(USUUS),T247   His| records | conflict with

                                    -3    +5 0   -4   -5-5 -2    +4
                                    his | permit in | several | ways.

                                    +5 -4   +5    0     -3     +5
PSS  X21-(SUSU)U(SSU)(UUSUUS),T248  Former convicts | are | pr |ime

                                    +5   0     -4    -5   +5  -1  -2   +3
                                    suspects | for the | hijacking case.
```

SUBSET 3D. Verb/Noun in Pairs (cont.)

PSS X22-(USSUUS)(US)U(USUU)U(US)(US),T249

-3 +4 | +3 -5 0 |
The news | mag | azine |

+5 | -2 | +5 -2 -3 |
Crime | sus | pects that the |

+5 -1 -3 -3 | -2 | +4 |
hijackers will | con | vict |

0 +5
them | selves.

PSS X23-S(US)(USU)(UUSU), T250

+5 0 +5 -2 | +5 0 | -5
Let's* record our | progress | to

-4 | +4 -3
the | tower.

PSS X24-S(US)(SSUSUU), T251

+5 -3 | +5 0 | +5 -2
Let's pro | gress towards | record

+5 -4 0
al | titude.

PSS X25-U U (SU)(SUS)(UUS)(USU)(UUS)(SSUU), T252

-2 | -2 +5 -1 |
Did | you increase |

+3 -4 | +4 -3 | -1 | +3 -5
record | l | ength to | per | mit the

+5 -1 -5 -7 +1 -2 |
program to pro gress mor | e

+5 -4 | -1
rapid | ly?

PSS X26-U U (SU)(USU)(USUSU)(USU)(UUSU)(UUSU)(SU), T253

-2 | +3 -3 +4 | -5 +4 | 0
Did | you record | the incre | ase

-5 | +4 -5 +5 | 0
in | rate of pro | gress

-4 | +5 | -2 -3 -5 | +4 -1
re | sul | ting from the | permit

-4 | -5 +4 -3 | +5 0
the | department | issued?

SUBSET 3D. Verb/Noun in Pairs

PSS X28-(SUSS)(USU)U(SU)S U S U U (USU)(SU)(USUUU), T255

-1 -5 | +5 | +3 -5 | +5 -4 -2
With a | strong | hand, the | lawmen were |

+3 -2 | +4 | -2 +4 | -1
ruling | Main | e, | but still | they

-3 | -2 +4 -2 | +5 0 | -4 +5 -4 -5 -2
were | recording | conflic | ts occasionally.

PSS X29-(USUUSS)(USU) U (SUS) U (SU) U U (USUSU)(UUS), T256

-3 +5 -3 -3 | +3 | -5 -4 | -5
According to | some | sources | the |

+5 -1 -2 | +5 -2 +2 | -2 +5 -2
lawmen were | lying men | but clearly

-2 -2 | -5 | +4 -1 | +5 -2
there were | con | flicting | records |

-4 +5 | +5
in the | files.

SUBSET 3F. Phonetic Influences on Simple Sentence Structures.

			Stress	Sentence
PSS	X253	SONOR,DECL1	+5 -1 +3 +5	Ron may* know* May.
PSS	X254	FRIC,DECL1	+5 -2 +3 +5	Sue has seen Fay.
PSS	X255	STOP,DECL1	+5 -4 +3 +5	Pete can take Kay.
PSS	X256	STOP,COMM1	+5 +5	Take Kay.
PSS	X257	STOP, COMM2	+4 +3 +5	Take* Kay* pop.
PSS	X258	STOP,COMM3	+5 +5 +4	Take* Kay too.
PSS	X259	FRIC,COMM1	+5 +4 +5	Serve* Sue fish.
PSS	X260	FRIC, COMM2	+5 +3 +5	Show* Sue how.
PSS	X261	STOP,Y-N1	-3 +5 +2 +5	Can Pete take Kay?
PSS	X262	STOP,Y-N2	-1 +5 +4 +5	Can Pete type too?
PSS	X263	FRIC,Y-N2	-2 +5 +3 +5	Has Sue* seen Fay?
PSS	X264	FRIC,Y-N2	0 +5 +2 +4	Has Sue* seen how?
PSS	X265	STOP,WH1	+5 -3 +4 +5	Who can take Kay?
PSS	X264	STOP,WH2	+5 -2 +4 +5	Where can Pete* Pack?
PSS	X267	FRIC,WH1	+4 -1 +4 +5	Who has* seen Fay?

SUBSET 4C. NP-PP-PP Subordination

PSS L262-SU(US)(SUSS)(US),T227

+3 -1 -5 +3 +5 -3 +3 +4
Ron will en* roll | many young men

-5 +5
in*May.

PSS L263-SU(US)(SUSS)(US),T228

+5 -1 -5 +3 +5 -3 +3
Ron will en*roll | many young

+4 -3 +5
men from*Maine.

PSS L264-SU(US)(SUSS)(US)(US,T229

+5 -1 -5 +3 +5 -4 +3
Ron will en | roll | many young

+4 -3 +5 -2 +4
men from | Maine | in May.

PSS L265-SU(US)(SUSS)(UUSUS),T227

+5 -1 -5 +3 -4 -4 +3
Ron will en*roll | many young

+5 -3 -4 +5 -4 +5
men in the | month of | May.

PSS L266-SU(US)(SUSS)(UUSU)(US),T230

+5 -1 -5 +3 +5 -4 +3
Ron will en | roll | many young

+5 -1 -4 -3 +5 -1 -4 +5
men | into the | army in | May.

PSS L267-SU(US)(UUSU)UU(UUS),T231

+1 -5 +5 -2 -5 +5 -3
Put the block on the | table

0 -5 +2 -5 +5
| which is by the | door.

PSS L268-(US)SS(UUSU)(UUUUS),T232

+1 -5 +5 -1 -5 -2 -5
Put the blo | ck which is | on the |

+5 -2 +4 -4 -1 -5 +5
table | ov | er by the | door.

SUBSET 6A. Commands.

PSS	S259 - S S (US),T21	+5 +4 -2 +5 Run mine a lone.
PSS	S260 - S S (SU),T21	+5 +4 +5 -1 Run mine early.
PSS	S261 - S S (SUU),T21	+5 +4 +5-4 +1 Run mine anyway.
PSS	S262 - S S (SUU),T21	+5 +5 +4-4 +1 Warn Ron anyway.
PSS	S263 - S(US) S,T21	+5 -3 +5 +4 Warn Ma rie now.
PSS	S264 - S (SU) S,T21	+5 +4 -2 +5 Warn Murray* now.
PSS	S275 - S (SUU) S,T21	+5 +5-5-2 +3 Warn Marion* now.
PSS	M55 - S (USU) S,T21	+5 -3+4-3 +5 Warn Ma ria now.
PSS	M56 - S (UUSU) S, T21	+5 +4-4 +4-3 +5 Warn Leo nora now.
PSS	M65 - S U (SU), T22	+5 -1 +4 -3 Loan me* money.
PSS	M66 - S (US) U, T22	+5 -4 +4 0 Loan Ma rie one.
PSS	M67 - S (SU) U, T22	+5 +5 -3 +1 L oan* Mary one.
PSS	M68 - S (US)(SU), T22	+5 -2 +3 +5 -1 Loan Ma rie money.
PSS	M69 - S (SU)(SU), T22	+5 +5 -2 +4 -1 Loan* Mary money.
PSS	M70 - S S (USUU), T22	+5 +4 -4 +5 -3 0 Loan* Ron a luminum.
PSS	M71 - S S (UUSUUS), T22	+5 +4 -2 -5 +5 -4-1 +4 Loan Ron an a luminum wire.
PSS	M72 -S (USU) S, T22	+5 -3 +4 -3 +4 Loan an airman rum.

SUBSET 6A. Commands (Cont.)

PSS M73 - S (UUSUU) S, T22	+5 -5 -2 +4-4-2 +5 Loan an Ar |menian| rum.
PSS M74 - S (UUSUUSU) S, T22	+5 -5 -1 +5-4-2 +4 -3 +5 Loan an Ar |menian| airman| rum.

SUBSET 7B. Yes/No Questions

PSS M143 - U (SS) S, T28

-1 +5 -4 +4
Will* your men kn ow?

PSS M144 - U (SSS) S, T28

-2 +5 -3 -2 +4
Will* all your men* know?

PSS M145 - U (SSSSU) S,T28

-3 +5 -3 +4 +5 -3 +4
Will | all your | nine airmen | know?

PSS M153 - U S (US)(SS),T30

-3 +5 -4 +5 +2 +5
Will | Ron* | en roll young | men?

PSS M154 - U S (US) (SUSS),T30

-3 +5 -3 +4 +5 -4 +1 +4
Will | Ron* en *roll many young me n?

PSS M155 - U S (US)(SUSSU),T30

-2 +5 -3 +4 +5 -2 +3 +5 -2
Will | Ron* en roll | really young air | men?

PSS M156 - U S (US)(SUSUUSUSS), T30

-2 +5 -4 +4 +5 -3 +4 -2 -3 +5 -4 +3 +3
Will | Ron en* roll | really | willing im | moral young | men?

PSS M159 - U U S S (SS),T32

-4 -3 +5 +4 -2 +4
Will I | owe* Ron my | ring?

PSS M160 - U U S S (SSSU),T32

-5 -2 +5 +4 +5 -1 +4 -3
Will I | owe* Ron | all my mon | ey?

PSS M163 - U U S (SUSU) S, T32

-4 -2 +5 +5-3 +4 -3 +5
Will I | owe | many | airmen ru m?

SUBSET 7D. W. H. QUESTIONS

PSS L1 - S U (US) S, T39
+5 -1 -4 +5 +4
When will Ma|rie* know?

PSS L2 - S U (SU) S, T39
+5 -1 +4 -3 +5
When will* Mary* know?

PSS L3 - S U (US)(US), T39
+5 +1 -4 +4 +4 +5
When will Ma|rie* en roll?

PSS L4 - S U (US)(SU), T39
+5 -1 -4 +4 +5 -3
When will Ma|rie* marry?

PSS L14 - S US (US) S, T40
+4 -1 +3 -4 +5 +4
When will |Ron* en roll* men?

PSS L15 - S US (US)(SS),T40
+4 -1 +5 -4 +4 +3 +5
When will |Ron* en roll young|men?

PSS L16 - S U S (US)(SUSSU),T40
+4 -1 +5 -5 +3 +3 -4
When will| Ron en|roll |really
+3 +5 -3
|young airmen?

PSS L27 - S S S (US),T98
+5 +3 +4 -4 +5
Who* loaned |rum to| Ron?

PSS L28 - S U S S (US), T101
+5 -2 +5 +4 -3 +4
Who will| warn| Ron i|n May?

PSS L36 - S U S S (US), T110
+5 -2 +4 +3 +5 0 -4 +5
Who will| Ron loan |rum to in| May?

PSS L37 - S U S S (US), T112
+5 -2 +4 +4 -4 +5
What will| men |loan to| Ron?

PSS L39 - S U S S (US) S, T114
+5 -2 +5 +5 -4 +4 +5
What| will| men |loan to |Ron |now?

PSS L40 - S U S S S (US), T115
+5 -2 +5 +3 +3 -3 +5
What| will |men |loan* Ron in |May?

PSS L41 - S U S S (US)(US), T116
+5 -2 +5 +4 -4 +5 -3 +5
What will| men |loan to|Ron in| May?

PSS L42 - S U S S S (US), T117
+5 -2 +5 +4 +4 -4 +5
When will| men* loan* rum| to Ron?

SUBSET 8A. Coordinate Sentences.

PSS L159 - S S S U S S S, T182

+5 +2 +5 -3 +5 +2
Ron* knew* Lynn | and Lou* knew*

+5
Wayne.

PSS L160 - S S S U S S S, T182

+5 +1 +5 -1 +5 +1
Ron* knew* Ly|nn or | Lou* knew |

+5
Wayne.

PSS L161 - S S S U S S S, T182

+4 +2 +5 -3 +4 +2
Ron | knew* Ly|nn and | Lynn* knew |

+5
Wayne.

PSS L162 - S S S U S S S, T182

+4 0 +5 -1 +4 0
Ron* knew* Lynn | or Lynn* knew |

+5
Wayne.

PSS L163 - S S S U S S S, T182

+4 +2 +5 -3 +5
Ron* knew | Lynn | and | Lynn*

+2 +5
knew | Ron.

PSS L164 - S S S U S S S, T182

+4 +1 +5 -2 +5 +1
Ron* knew* Ly|nn or Lynn* knew |

+5
Ron.

PSS L165 - S S S U S S S, T182

+5 +1 +3 -3 +5 -1
Ron* knew | Lynn and | Lou* knew |

+3
Lynn.

PSS L166 - S S S U S S S, T182

+5 +2 +3 -2 +5 0
Ron* knew* Lyn|n |or Lou* knew*

+2
Lynn.

PSS L167 - S U S S S U S U S S, T183

+4 -3 -1 +5 -3 +4
Ron may* know | Lynn | and Lou |

-3 -1 +5
may* know | Wayne.

SUBSET 8A. Coordinate Sentences.

PSS L168 - S U S S U S U S S, T183

+4 -3 -1 +5 -3 +5
Ron may* know | Lynn | and Lynn

-3 -1 +5
may* know | Wayne.

PSS L169 - S U S S U S U S S, T183

+5 -4 -1 +5 -3 +5
Ron may* know | Lynn | and Lynn

-4 -1 +5
may* know | Ron.

PSS L170 - S U S S U S U S S, T183

+5 -3 +3 +3 -2 +4 -4
Ron may | know | Lynn or | Lynn may*

-1 +5
know | Ron.

PSS L171 - S U S S U S U (SU) S, T183

+4 -2 +5-4 +2 -2 +4 -2
Ron may | ruin | Lynn | or Ron may |

+5 -2 +2
marry* Lynn.

PSS L172 - S U S S U S U (SU) S, T183

+4 -1 +4 -5 +2 -4 +4 -1
Ron may | ruin | Lynn and* Ron may |

+4 -3 +2
marry* Lynn.

PSS L173 - S U S S U S U S S S, T183

+5 -1 +4 +1 -3 +3 -1
Ron may* know* Lynn | or Ron may

+5 -1 +1
| not* know* Lynn.

PSS L174 - S U S S U S U U (SU) S, T183

+5 -1 +5 -5 +2 -4 +3
Ron may | ruin | Lynn, but | Ron

-2 +5 +4 -3 +2
may | not | marry Lynn.

PSS L175 - S U U(SU)S U S U (SU) S, T183

+5 -2 -1 +5 -3 +2 -3
Ron may not | marry Ly|nn, but |

+4 -1 +5-3 +2
Ron may | ruin | Lynn.

PSS L176 - S U S S S U S U S S S, T184

+5 -3 +3 +5 +5 -3
Ron may | loan | Lynn | rum | and

+5 -2 +4 +5 +5
Lou may | loan | Wayne | oil.

SUBSET 8A. Coordinate Sentences.

PSS L177 - S U S S S (US) U S U S S S (US), T185

+5 -3 +1 | +4
Ron may* loan | Lynn

+5 -2 +4 -2 +4
| rum in | May | and Lou

-2 -1 +3 +2
may* loan | Wayne | oil

-3 +2
in | May.

PSS L178 - S U S S S (US) U S U S S S (US), T185

+5 -3 +3 +5
Ron may | loan | Lynn |

+4 -3 +4 -3 +5
rum in | May | and Lou

-4 0 +3 +1
may | loan | Wayne* rum

-4 +3
in* May.

SUBSET 8H. Coordinate Verb Phrases

PSS L206 - (SU) S S U S S, T129

+4 -5 +4 +3 -5 +3
Women | own y | arn and | wear |
+4
wool.

PSS L207 - S U (SU) S U (SU) S, T193

+5 -2 +5-4 +1 -5
R | on may | ruin | Lynn | and
+5 -2 0
marry* Lynn.

PSS L208 - S U (SU) S U (SU) S, T193

+4 -2 +5 -4 +1 -5
Ron may | ruin | Lynn | a nd
+5 -2 +5
marry | Lou.

PSS L209 - S U (SU) S (SU) S U (US) S, T194

+3 -2 +5 -4 +2 +4 -3
Ron may | ruin | Lynn, | marry
+4 -4 -4 +5 +5
Lo | u, and | annoy | Wayne.

PSS L210 - U (SU) S S U S S, T195

-1 +5 -4 +4 +4 -2
Will | women | own y | arn and |
+4 +5
wear | wool?

PSS L211 - U (SU) S S S S U S S, T196

-2 +5 -4 +5 +3 +4
Will women | own | yarn, | wear |
+3 -2 +5 +4
wool, and | woo | men?

PSS L212 - U S (SU) S (SU) S U (US) S, T196

-2 +4 +5-4 +4 +4 -2
Will | Ron | ruin | Lynn, | marry |
+4 -2 -3 +4 +4
Lou, and an | noy | Wayne?

PSS L213 - S S S U S S, T197

+3 +4 +3 -5 +4
Who | owns* yarn and | wears*
+4
wool?

PSS L214 - S S S U S S, T197

+2 +3 +4 -3 +4
Who* owns* yarn | or wears*
+4
wool?

SUBSET 8H. Coordinate Verb Phrases

PSS L215 - S U (SU) S U S (SU) S, T193

+5 -1 +5-5 +3 -5
Ron may | ruin | Lynn | and |

+2 +3 -4 +1
not | marry* Lynn.

PSS L216 - S U (SU) S U S (SU) S, T193

+4 -1 +5 -5 +2 -3 +3
Ron may | ruin | Lynn or not |

+5 -4 +1
marry* Lynn.

PSS L217 - S U (SU) S U S (SU) S, T193

+5 -1 +5 -5 +2 -3 +2
Ron may | ruin L | ynn but | not |

+5 -3 +1
marry* Lynn.

SUBSET 8K. Coordinate Noun Phrases

PSS L242 - (SUS) S (SUS) (USUU), T215

+5 -4 +5 -1 +5 -4
Lou and | Neal | knew | Ron and |

+5 -4 +5 -5 -2
Lynn |respectively.

PSS K243 - (SSUS) S (SSUS)(USUU), T216

+4 +4 -4 +5 -1
Wayne, | Lou, and |Neal | k |new

+4 +4 -4 +5
Lee, |Ron, |and | Lynn, |

-4 +4 -5 -2
respectively.

PSS L244 - (SUS) U S (SUS)(USUS), T217

+5 -4 +5 -4 0 +5
Lou and | Neal | will loan* oil

-4 +5 -2 +5 -3 +5
an |d ore to | Lynn and* Ron.

PSS L245 - U S (USS) U (USS), T202

-2 +2 -5 0 +5 -4 -5 -1 +4
I saw an | old house and an | old barn.

PSS L246 - U S (USS) U (USS), T202

-2 +1 -5 +5 +1 -4 -5 +5 0
I saw an | old | house and a | new house.

PSS L247 - (USUSUUS) U S (SUS) U (SUSUS), T202

-3 +5 -4 +5 -5 -3 +4
The two a |vaila |ble meals

+3 +5 +5 -4 +4
were | t |oast,* ham and eggs

-2 +5 +1 +5 -2 -4 +5
and pancakes,| syrup and | juice.

PSS L248 - (SSSUUS) U (SUS) U (SUUS), T202

-2 +2 +4 -5 -3 +5 -1
My two favorite | meals | are |

+4 -4 +5 -1 -4 -2 -5
ham and eggs | and |pizza and |

+5
beer.

PSS L249 - (SSSUUS) U S (SUS) U (SU), T206

-3 +4 +4 -5 -3 +5 -1
My three favorite | meals are |

+5 +4 -5 +5 -2
hash, | ham and eggs and |

+5 -2
pizza.

SUBSET 8K. Coordinate Noun Phrases

PSS L250 - (SSSUUS) U (SUS)(SUS) U (SUUS), T206

-3 +4 +4-5 -3 +5 -2
My three favorite | meals | are

+4 -4 +3 +4 -4 +3
| hash and | beans, ham and eggs |

-4 +5 -2 -3 +4
and | pizza and | beer.

SUBSET 11A. Relative Clauses

PSS L107 - S U S S S S, T142

+5 -1 +4 +4 +3 +5
Men | who | knew* Ron | ran* Maine.

PSS L108 - S U S S S S, T143

+5 -2 +5 -1 +2 +5
Men | whom Ron* knew | ran | Maine.

PSS L109 - (USU) U (US) S S S, T143

-3 +5 -4 -1 -3 +3 0 +3 +5
The| a|irmen |whom |Marie* knew | ran* Maine.

PSS L110 - S S S S S, T143

+4 +4 +4 +1 +5
Men |Ron* knew | ran Maine.

PSS L111 - S U S S S S, T144

+5 -2 +2 +5 +2 +5
Lyn |n, |who |knew* Ron, |ran* Maine.

PSS L112 - S U S S S S, T145

+5 -3 +5 0 +3 +5
Lynn, | w |hom| Ron* knew, | ran | Maine.

PSS L113 - S U S S (US) S S, T146

+4 -2 +5 +2 -5 +5 +4 -4 +5
Oil which| Ron* Loaned to |Ann | ruined* Maine.

PSS L114 - S S S (US) S S, T146

+4 +4 +3 -2 +3 +3 -3 +4
Oil| Ron* loaned to* Ann| ruined* Maine.

PSS L115 - S U U U S U S S, T147

+5 -1 0 -1 +5 -1 +2 +5
|Men |who are in |Rome |may |run| Maine.

PSS L116 - S U S U S S, T147

+4 -3 +4 -2 +2 +5
Men in* Rome|may|run* Maine.

PSS L117 - S S S S S, T152

+5 +1 +2 +3 +5
Men,* Ron* knew |Ran* Maine.

PSS L118 - S S S U S S, T148

+4 +1 +3 -4 +5 +1
Ann* knew* men |whom |Ron* knew.

SUBSET 11A. Relative Clauses

PSS L119 - S S S S S, T148

+5 +3 +3 +5 +1
Ann* knew* men* Ron* knew.

PSS L120 - S S S U S S, T149

+5 +2 +5 -2 +5 +1
Ann* knew* May, |whom Ron* knew.

PSS L121 - S S S U S S, T149

+5 +2 +5 -3 +5 +1 +5
Ann* knew* May |whom |Ron knew |too.

PSS L122 - S S S (US) U S S, T150

+5 +4 +5 -4 +5 -2 +5 +4
Ron* loaned | Lynn |the |oil | which| Wayne* owned.

PSS L123 - S S S (US) S S, T150

+5 +4 +5 -4 +4 +4 +5
Ron* loaned |Lynn the |oil |Wayne* owned.

PSS L124 - S S S (US) U S (US), T151

+5 +3 +4 -3 +5 -2 +4 -2 +4
Ron* loaned |Lynn the |oil,| which |Wayne be|moaned.

APPENDIX B: PRESENTATIONS AND PUBLICATIONS BY SPERRY UNIVAC DURING THE ARPA/SUR PROGRAM

The following is a complete list of Sperry Univac publications, reports, presentations, and unpublished ARPA SUR Notes, resulting from ARPA funding during the SUR program:

Publications and Reports

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E., Prosodic Aids to Speech Recognition: I. Basic Algorithms and Stress Studies, Univac Report No. PX 7940, Univac Park, St. Paul, Minnesota, October 1972.

LEA, W. A., Influences of Phonetic Sequences and Stress on Fundamental Frequency Contours of Isolated Words, J. Acoust. Soc. of America, Vol. 53, January, 1973, 346(A).

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E., Use of Syntactic Segmentation and Stressed Syllable Location in Phonemic Recognition, J. Acoust. Soc. of America, Vol. 53, January, 1973, 356(A).

LEA, W.A., Syntactic Boundaries and Stress Patterns in Spoken English Texts, Univac Report No. PX 10146, Univac Park, St. Paul, Minnesota, March, 1973.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E., Prosodic Aids to Speech Recognition: II. Syntactic Segmentation and Stressed Syllable Location, Univac Report No. PX 10232, Univac Park, St. Paul, Minnesota, April, 1973.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E., Prosodic Aids to Speech Recognition: III. Relationships between Stress and Phonemic Recognition Results, Univac Report No. PX 10430, Univac Park, St. Paul, Minnesota, September, 1973.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E., A Prosodically-Guided Speech Understanding Strategy, Proc. IEEE Symposium on Speech Recognition, Carnegie-Mellon University, Pittsburgh, Penn., April, 1974, 38-44.

LEA, W. A. Prosodic Aids to Speech Recognition IV: A General Strategy for Prosodically-Guided Speech Understanding, Univac Report No. PX 10791, Univac Park, St. Paul, Minnesota, March, 1974.

LEA, W. A., Sentences for Controlled Testing of Acoustic Phonetic Components of Speech Understanding Systems, Univac Report PX 10952, Univac Park, St. Paul, Minnesota, September, 1974.

LEA, W. A., Prosodic Aids to Speech Recognition. V. A Summary of Results to Date, Univac Report No. PX 11087, Univac Park, St. Paul, Minnesota, October, 1974.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E., A Prosodically-Guided Speech Understanding Strategy, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-23, February, 1975, 30-38.

LEA, W. A. and KLOKER, D. R., Prosodic Aids to Speech Recognition: VI. Timing Cues to Linguistic Structure and Improved Computer Programs for Prosodic Analysis, Univac Report No. PX 11239, Univac Park, St. Paul, Minnesota, March, 1975.

LEA, W. A., Prosodic Aids to Speech Recognition: VII. Experiments on Detecting and Locating Phrase Boundaries, Univac Report No. PX 11534, Univac Park, St. Paul, Minnesota, November, 1975.

LEA, W. A., Acoustic Correlates of Stress and Juncture, Univac Report No. PX 11693 Univac Park, St. Paul, Minnesota, June, 1976. To appear in Stress and Accent (L. Hyman, Ed.), University of Southern California Press, Los Angeles.

LEA, W. A., The Importance of Prosodic Analysis in Speech Understanding Systems, Univac Report No. PX 11694, Univac Park, St. Paul, Minnesota, June, 1976, Submitted to IEEE Trans. Acoustics, Speech and Signal Processing.

LEA, W. A., Prosodic Aids to Speech Recognition: VIII. Listeners' Perceptions of Selected English Stress Patterns, Univac Report No. PX 11711, Univac Park, St. Paul, Minnesota, June, 1976.

LEA, W. A., Sentences and Hypotheses for Controlled Testing of Syntactic and Prosodic Components of Speech Understanding Systems, Univac Report No. PX 10953 Univac Park, St. Paul, Minnesota, November, 1976.

LEA, W. A. Prosodic Aids to Speech Recognition: IX. Acoustic-Prosodic Patterns in Selected English Phrase Structures, Univac Report No. PX 11963, Univac Park, St. Paul, Minnesota, December, 1976.

Oral Presentations

KLOKER, D. R. (April 1975). "Vowel and Sonorant Lengthening as Cues to Phonological Phrase Boundaries." presented at the 89th Meeting of the Acoustical Society of America, Austin, Texas.

KLOKER, D. R. (April 1976). "A Technique for the Automatic Location and Description of Pitch Contours," presented at the 1976 International Conference on Acoustics, Speech and Signal Processing, Philadelphia, Pennsylvania.

LEA, W. A., Influences of Phonetic Sequences and Stress on Fundamental Frequency Contours of Isolated Words, presented at the 84th Meeting of the Acoustical Society of America, Miami Beach, Florida, November, 1972

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (November 1972). "Use of Syntactic Segmentation and Stressed Syllable Location in Phonemic Recognition," presented at 84th Meeting of the Acoustical Society of America, Miami Beach, Florida.

LEA, W. A., Prosodic Features and Linguistic Structure, presented at the ARPA Tutorial Lectures on Acoustic-Phonetic Characteristics of English Sentences, Cambridge, Massachusetts, December, 1972.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E., A Prosodically Guided Speech Understanding Strategy, presented at IEEE Symposium on Speech Recognition, Carnegie-Mellon University, Pittsburgh, PA., April, 1974.

LEA, W. A., "Perceived Stress as the 'Standard' for Judging Acoustical Correlates of Stress", presented at the 86th Meeting of the Acoustical Society of America, Los Angeles, California, November, 1973.

LEA, W. A., "Evidence that Stressed Syllables Are the Most Readily Decoded Portions of Continuous Speech", presented at the 86th Meeting of the Acoustical Society of America, Los Angeles, California, November, 1973.

LEA, W. A., "An Algorithm for Locating Stressed Syllables in Continuous Speech", presented at the 86th Meeting of the Acoustical Society of America, Los Angeles, California, November, 1973.

LEA, W. A., Prosodic Phenomena, session chaired at ARPA Phonological Rules Workshop, Systems Development Corporation, Santa Monica, California, June, 1974.

LEA, W. A., A Speech Data Base for Testing Components of Speech Understanding Systems, presented at 88th Meeting of the Acoustical Society of America, St. Louis, Missouri, November, 1974.

LEA, W. A., Prosodic Hypotheses and Rules, presented at the ARPA Workshop on Acoustic Phonetic and Phonological Rules, Bolt Beranek and Newman, Cambridge, Massachusetts, November, 1974.

LEA, W. A., Isochrony and Disjuncture as Aids to Syntactic and Phonological Analysis, presented at the 89th Meeting of the Acoustical Society of America, Austin, Texas, April, 1975. Abstract in J. Acoust. Soc. America, Vol. 57, Suppl. No. 1, Spring, 1975.

LEA, W. A., Acoustic Correlates of Stress and Juncture: A systematic Testing of Alternative Hypotheses, presented at the Symposium on Stress and Accent, University of Southern California, February, 1976.

LEA, W. A., Stress om English: Listeners' Perceptions and Acoustic Correlates, presented to the Linguistics Club, University of Minnesota, Minneapolis, May, 1976.

LEA, W. A., Perceived Stress Patterns in Selected English Phrase Structures, presented to the American Assoc. of Phonetic Sciences, San Diego, California, November 15, 1976.

LEA, W. A., Use of Intonational Phrase Boundaries to Select Syntactic Hypotheses in a Speech Understanding System, presented at the 92nd Meeting, Acoustical Society of America, San Diego, California, November 16, 1976. J. Acoust. Soc. of America, vol. 60, Suppl. 1, Page S12.

Unpublished ARPA SUR Notes

2. MEDRESS, M. F., The Univac Speech Recognition Study (5 pages), December, 1971.

16. MEDRESS, M. F., Proposed Computer Phonetic Transcriptions (2 pages), February, 1972.

17. MEDRESS, M. F., Univac Speech Bibliography (1 page), February, 1972.

32. MEDRESS, M. F., Revised Computer Phonetic Representations (2 pages), April, 1972.

36. MEDRESS, M.F., et al., "Acoustic Correlates of Linguistic Stress" (Literature Survey, (22 pages)), June, 1972.

39. LEA, W. A., Acoustic Cues for Boundaries between Syntactic Units (6 pages), August, 1972.

45. LEA, W. A., Considerations in the Design of Good Speech Texts (7 pages), September, 1972.

48. MEDRESS, M.F., Plans for the First Segment of the Speech Data Base (2 pages), October, 1972.

53. LEA, W. A., MEDRESS, M. F., and SKINNER, T.E., Use of Syntactic Segmentation and Stressed Syllable Location in Phonemic Recognition (11 pages), December, 1972.

54. LEA, W. A., Syntactic Factors in the Initial Selection of Sentences for the Data Base (7 pages), December, 1972.

63. LEA, W. A., Some Factors in the Selection of Utterances for Speaker Normalization (3 pages), February, 1973.

67. LEA, W. A., Acoustic Analysis of 13 ARPA Sentences (4 pages), February, 1973.

82. LEA, W. A., MEDRESS, M. F., and SKINNER, T.E., Univac Final Technical Report: Prosodic Aids to Speech RecognitionII: Syntactic Segmentation and Stressed Syllable Location (34 pages), May, 1973.

108. LEA, W. A., MEDRESS, M. F., and SKINNER, T.E., Prosodic Aids to Speech Recognition: III. Relationships between Stress and Phonemic Recognition Results (6 pages), October, 1973.

139. MEDRESS, M. F., Prosodic Aids to Speech Recognition: IV, A General Strategy for Prosodically-Guided Speech Understanding (65 pages), May, 1974.

141. LEA, W. A., Sentences for Testing Acoustic Phonetic Components of Systems (18 pages), July, 1974.

154. LEA, W. A., Sentences for Controlled Testing of Acoustic Phonetic Components of Speech Understanding Systems (41 pages), November, 1974.

155. LEA, W. A., Sentences for Testing Prosodic and Syntactic Components of Systems (52 pages), November, 1974.

158. LEA, W. A., Prosodic Aids to Speech Recognition: V. A Summary of Results to Data (29 pages), January, 1975.

162. LEA, W. A., An Improved Program for Detecting Boundaries Between Syntactic Constituents (6 pages), March, 1975.

163. LEA, W. A., A Computer Program for Locating Stressed Syllables (27 pages), March, 1975.

182. LEA, W. A., Prosodic Algorithms and Studies of Timing Cues to Linguistic Structure (59 pages), August, 1975.

197. LEA, W. A., Experiments on Detecting and Locating Phrase Boundaries, (43 pages), January, 1976.

206. LEA, W. A., Two Papers about Prosodic Structures (131 pages), June, 1976.

213. LEA, W. A., Two Studies: Listeners' Perceptions of Stess and Prosodic Aids to the BBN Parser (70 pages), July, 1976.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Univac Defense Systems Division
P. O. Box 3525
St. Paul, Minnesota 55165

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

(6) Prosodic Aids to Speech Recognition: IX. Acoustic Prosodic Patterns in Selected English Phrase Structures.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

(9) Final Technical Report. 1 Sept[redacted] 1975 - 30 Nov[redacted] 1976,

5. AUTHOR(S) (First name, middle initial, last name)

(10) Wayne A. Lea

(12) 116p.

6. REPORT DATE

(11) 31 Dec[redacted] 1976

7a. TOTAL NO. OF PAGES

105 + vii

7b. NO. OF REFS

36

8a. CONTRACT OR GRANT NO.

(15) DAHC 15-73-C-0310

b. PROJECT NO.

ARPA Order-2010

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

(14) Univac Report No. PX-11963

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

13. ABSTRACT

The final two studies have been completed in a four year effort on developing prosodic aids to speech recognition. A procedure for using intonational phrase boundaries to select among alternative word and phrase hypotheses has been developed, refined, and tested by hand analyses of sixteen sentences. This procedure was designed for use with the BBN HWIM speech understanding system, and was totally implemented, but not tested before the end of the BBN contract with ARPA. Comparisons of control and parsing traces with acoustically detected phrase boundaries did show, however, that intonational boundaries could help select correct words and phrase structures and avoid erroneous hypotheses.

An experimental study of acoustic prosodic patterns in 255 sentences showed several useful prosodic regularities. Over 91% of the syllables were correctly located, and 92% of the stressed syllables were correctly categorized as stressed, while 76% of the syntactic phrase boundaries were detected. Exactly which phrases are or are not preceded by intonational phrase boundaries was determined. Intonation contours were very firmly shown to involve rising pitch until the first stress, progressively lower pitch in succeeding stresses, and a terminal fall (for declaratives, commands, and WH questions) or rise (for yes/no questions).Parantheticals were clearly marked by disjunctures, large Fo variations, and other prosodic features. Contrastive phrase structures could be detected from prosodic cues.

A summary of Sperry Univac's total contributions to ARPA/SUR shows efforts to define the importance of prosodics in speech understanding, to cooperate with other contractors, and to conduct experiments on all aspects of prosodic structure. Further work is suggested.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

LB

408125

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Recognition						
Speech Analysis						
Linguistic Stress						
Prosodies						
Prosodic Features Extraction						
Intonation						
Syntactic Boundary Detection						
Stressed Syllable Location						
Syntactic Analysis						
Syntactic Parsing						

Security Classification